



GAUSSIAN MIXTURE REDUCTION FOR
BAYESIAN TARGET TRACKING IN CLUTTER

THESIS

David J. Petrucci, Captain, USAF

AFIT/GE/ENG/06-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, United States Department of Defense, or the United States Government.

AFIT/GE/ENG/06-01

GAUSSIAN MIXTURE REDUCTION FOR
BAYESIAN TARGET TRACKING IN CLUTTER

THESIS

Presented to the Faculty
Department of Electrical and Computer Engineering
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Electrical Engineering

David J. Petrucci, B.S.E.E.

Captain, USAF

22 December 2005

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GE/ENG/06-01

GAUSSIAN MIXTURE REDUCTION FOR
BAYESIAN TARGET TRACKING IN CLUTTER

David J. Petrucci, B.S.E.E.
Captain, USAF

Approved:

/signed/

9 Nov 2005

Dr. Peter S. Maybeck (Chairman)

date

/signed/

9 Nov 2005

Lt Col Juan R. Vasquez (Member)

date

/signed/

9 Nov 2005

Dr. Richard K. Martin (Member)

date

Abstract

The Bayesian solution for tracking a target in clutter results naturally in a target state Gaussian mixture probability density function (pdf) which is a sum of weighted Gaussian pdfs, or mixture components. As new tracking measurements are received, the number of mixture components increases without bound, and eventually a reduced-component approximation of the original Gaussian mixture pdf is necessary to evaluate the target state pdf efficiently while maintaining good tracking performance. Many approximation methods exist, but these methods are either *ad hoc* or use rather crude approximation techniques. Recent studies have shown that a measure-function-based mixture reduction algorithm (MRA) may be used to generate a high-quality reduced-component approximation to the original target state Gaussian mixture pdf.

To date, the Integral Square Error (ISE) cost-function-based MRA has been shown to provide better tracking performance than any previously published Bayesian tracking in heavy clutter algorithm. Research conducted for this thesis has led to the development of a new measure function, the Correlation Measure (CM), which gauges the similarity between a full- and reduced-component Gaussian mixture pdf. This new measure function is implemented in an MRA and tested in a simulated scenario of a single target in heavy clutter. Results indicate that the CM MRA provides slightly better performance than the ISE cost-function-based MRA, but only by a small margin.

Acknowledgments

This thesis marks the end and successful completion of over four years of part-and full-time graduate education. Over those years, many organizations and individuals provided invaluable advice and support to my endeavor. The Dayton Area Graduate Studies Institute financially supported a large portion of the educational expenses for my part-time graduate education, and the management from my previous assignment provided the flexibility in duty hours to attend classes. Dedicated graduate instructors at both institutions went beyond teaching problem-specific equations and emphasized over-arching concepts applicable to a wide variety of problems.

In particular, there are several individuals who deserve special recognition. Through his cutting-edge courses on estimation, multiple target tracking, and multi-sensor data fusion, Dr. Lang Hong was the sole influence for my choice of specialty. Since the fall of 2002, Dr. Peter Maybeck has positively influenced both professional and academic aspects of my career, the impact of which will surely pervade throughout my time in the USAF and beyond. His unique combination of unsurpassed knowledge of the specialty and genuine interest in the student's academic success make Dr. Maybeck the ideal thesis advisor. Insightful feedback provided by Lt Col Vasquez and Dr. Martin was greatly appreciated. I especially would like to thank Jason Williams, whose ground-breaking research in the development of his ISE cost-function-based tracking filter provided the basis for this thesis.

Finally, I leave the reader with an interesting quotation which describes an experience that I believe all of us have had during our graduate studies (hopefully!).

“Everyone has come across the sort of problem which seems impossible to solve until suddenly a surprisingly simple solution is revealed. Once it has been thought of, the solution is so obvious that one cannot understand why it was ever so difficult to find.”

Edward de Bono, *The Use of Lateral Thinking* (Penguin Books, 1977)

David J. Petrucci

Table of Contents

	Page
Abstract	iv
Acknowledgments	v
List of Figures	ix
List of Tables	xi
List of Abbreviations	xii
Notation	xiii
I. Introduction	
1.1 Research Goal	3
1.2 Organization	3
II. The Bayesian Approach to Target Tracking	
2.1 Target Kinematics Models	8
2.2 Recursive Bayesian Filtering	12
2.2.1 Linear Recursive Bayesian Filtering	12
2.2.2 Nonlinear Recursive Bayesian Filtering	19
2.3 Multivariate Gaussian Mixtures	26
2.4 Bayesian Approaches for Kinematics Model Parameter Uncertainty	29
2.4.1 Non-Switching Models	30
2.4.2 Switching Models	38
2.4.3 Multiple Model Algorithms Summary	57
2.5 A Bayesian Approach for Measurement Origin Uncertainty	59
2.5.1 A Bayesian Solution for Measurement Origin Uncertainty	65
2.5.2 Tracking with Measurement Origin Uncertainty Summary	78
2.6 Summary	79
III. Estimating Probability Density Functions	
3.1 Maximum Likelihood Estimation	82
3.1.1 Asymptotic Properties of MLE	83
3.1.2 MLE Measure Function	90
3.2 Expectation Maximization	91

	Page	
3.2.1	Theoretical Derivation of the EM Algorithm	92
3.2.2	EM Algorithm Example	96
3.2.3	General Form of the EM Algorithm	99
3.3	Multivariate Gaussian Mixture Estimation	100
3.3.1	MLE of a Multivariate Gaussian Mixture	102
3.3.2	EM Algorithm for a Multivariate Gaussian Mix- ture	104
3.3.3	Asymptotic Representation of the EM Algorithm	106
3.4	Summary	107
IV.	Approximating Gaussian Mixtures & Mixture Reduction Algorithms	108
4.1	Measure Functions for Gaussian Mixture Approximation	109
4.1.1	True Distance Measures	111
4.1.2	Pseudo-Distance Measures	119
4.2	Greedy Algorithms for the Assignment Problem	121
4.3	Salmond's Joining & Clustering Algorithms	122
4.4	Williams' ISE Cost-Function-Based Algorithm	124
4.5	Summary	127
V.	Gaussian Mixture Reduction Algorithm Development & Analysis	128
5.1	Measure Function & Assignment Algorithm Selection . .	128
5.2	Mixture Reduction Algorithm Development	132
5.2.1	Closed-Form Solutions of Select Measure Func- tions	132
5.2.2	Proposing Mixture Reduction Actions	140
5.2.3	Mating Measure Functions with Assignment Al- gorithms	141
5.3	Mixture Reduction Algorithm Analysis	150
5.3.1	Impact of Mixture Weight Re-Normalization . .	151
5.3.2	Mixture Reduction Algorithm Test Results . . .	155
5.4	Summary	158
VI.	Simulation Description & Results	159
6.1	Description	159
6.2	Results	161
6.3	Summary	165
VII.	Conclusions & Recommendations	166
7.1	Restatement of the Research Goal	166
7.2	Summary of Results	166
7.3	Significant Contributions of Research	169
7.4	Recommendations for Future Research	170

Bibliography	172
------------------------	-----

List of Figures

Figure		Page
2.1	Target Tracking Block Diagram.	7
2.2	Kalman Filter Effects on the Conditional Target State pdf.	19
2.3	A Univariate Gaussian Mixture pdf.	27
2.4	Non-Switching Model Block Diagram.	37
2.5	GPB-1 Block Diagram.	46
2.6	GPB-2 Block Diagram.	52
2.7	IMM Block Diagram.	57
2.8	The Data Association Problem.	60
2.9	Measurement Gating.	61
2.10	Generated Association Events.	63
3.1	Sample Spaces for the EM Algorithm Derivation.	92
4.1	Vectors Representing pdfs in Hilbert Space.	110
4.2	ISE Cost Function-based Algorithm Flowchart.	126
5.1	Two Distinct Univariate Gaussian Mixture pdfs.	130
5.2	A Flowchart of the Correlation Measure Mixture Reduction Algorithm.	142
5.3	A Flowchart of the ISE Shotgun Mixture Reduction Algorithm.	145
5.4	A Flowchart of the Hellinger Distance Mixture Reduction Algorithm.	146
5.5	A Flowchart of the Hellinger Affinity Measure Mixture Reduction Algorithm.	149
5.6	Two Sets of Vectors in Hilbert Space.	152
5.7	Illustration of the Effect of Mixture Weight Re-Normalization.	154
5.8	Results from the First Mixture Reduction Algorithm Test.	156
5.9	Results from the Second Mixture Reduction Algorithm Test.	157

Figure		Page
6.1	The First Comparison of the Track Life Results for the Correlation Measure and Integral Square Error Mixture Reduction Algorithms.	162
6.2	The Second Comparison of the Track Life Results for the Correlation Measure and Integral Square Error Mixture Reduction Algorithms.	162
6.3	The Third Comparison of the Track Life Results for the Correlation Measure and Integral Square Error Mixture Reduction Algorithms.	164
7.1	Various Vectors Representing Two pdfs in Hilbert Space.	169

List of Tables

Table		Page
5.1	Univariate Gaussian Mixture pdf Parameters.	130

List of Abbreviations

Abbreviation		Page
pdf	probability density function	1
PDAF	Probabilistic Data Association Filter	1
JPDAF	Joint Probabilistic Data Association Filter	1
ISE	Integral Square Error	2
MRA	mixture reduction algorithm	3
CV	constant velocity	11
EKF	Extended Kalman Filter	20
MMAE	Multiple Model Adaptive Estimation	31
pmf	probability mass function	32
GPB	Generalized pseudo-Bayesian	41
IMM	Interacting Multiple Model	41
MLE	maximum likelihood estimation/estimate	81
EM	Expectation Maximization	82
i.i.d.	independent identically distributed	82
CM	Correlation Measure	131
HD	Hellinger Distance	131
HA	Hellinger Affinity Measure	131

Notation

Notation	Usage
$\boldsymbol{x}(k)$	the discrete-time state random process vector at sample k ; represents the state (location, velocity, etc.) of the target
$\hat{\boldsymbol{x}}(k k-1)$	the mean estimate of the state random process vector at sample k , using information only up to the $(k-1)^{\text{th}}$ measurement
$\mathbf{P}(k k)$	the covariance estimate of the state random process vector at sample k , using all available information up to the k^{th} measurement
$\boldsymbol{X}(k)$	the joint target state random process composite vector containing the state random process vectors of multiple targets at sample k
$\boldsymbol{z}(k)$	the measurement random process vector at sample k
$\boldsymbol{Z}(k)$	the composite measurement random process vector containing all of the measurement random process vectors at sample k
\mathbf{z}_k	a realization of $\boldsymbol{z}(k)$ at sample k
\mathbf{Z}_k	a realization of $\boldsymbol{Z}(k)$ at sample k
\mathbf{Z}^k	the measurement history through sample k
\mathbf{Z}^k	the realized measurement history through sample k
$p(\cdot)$	the probability mass function (pmf) for the discrete random argument (\cdot)
$f(\cdot)$	the probability density function (pdf) for the continuous random argument (\cdot)
$L(\{z_i\}_1^n; \alpha)$	the likelihood function for the set of observations $\{z_i\}$, $i = 1, \dots, n$, from the distribution of z with pdf scalar parameter α unknown
$\mathcal{N}\{x; \mu, \sigma^2\}$	denotes a Gaussian pdf for the scalar random variable x , distributed with mean μ and variance σ^2
$\mathcal{N}\{\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P}\}$	denotes a Gaussian pdf for the vector random variable \boldsymbol{x} , distributed with mean $\boldsymbol{\mu}$ and covariance \mathbf{P}
$\Omega_o, \hat{\Omega}$	the set of multivariate Gaussian mixture parameters for the original and reduced-component target state pdfs, respectively
$\alpha_o, \hat{\alpha}$	the true and estimated pdf scalar parameters, respectively
N_f	the number of models used in a multiple model algorithm, and thus the number of elemental filters
\mathbf{M}	continuous random vector representing kinematic model uncertainty for non-switching models
$\mathbf{M}(k)$	continuous random process vector representing kinematic model uncertainty for switching models
\mathbf{M}_i	discrete random vector representing kinematic model uncertainty for non-switching models

Notation	Usage
\mathbf{M}_{i_k}	discrete random process vector representing kinematic model uncertainty for switching models
\mathbb{M}	the continuous sample space of \mathbf{M} and $\mathbf{M}(k)$; $\mathbf{M}, \mathbf{M}(k) \in \mathbb{M} \subset \mathbb{R}^n$; or, the discrete sample space of \mathbf{M}_i and \mathbf{M}_{i_k} ; $\mathbf{M}_i \in \mathbb{M} = \{\mathbf{M}_i\}_1^{N_f}$ and $\mathbf{M}_{i_k} \in \mathbb{M} = \{\mathbf{M}_i\}_1^{N_f}$
$\{\mathbf{M}_{i_\ell}\}_1^{k-1}$	the history of models for switching models from sample $\ell = 1$ through sample $\ell = k - 1$
$\mu_i(k), \mu_{i_k}(k)$	mode probability that model i for non-switching models, or model i_k for switching models, is correct given the measurement history
$\tau_{i_k, i_{k-1}}$	mode Markov transition probability (from model $\mathbf{M}_{i_{k-1}}$ to model \mathbf{M}_{i_k}) for switching models under the assumption that \mathbf{M}_{i_k} is a discrete Markov random process
$\Theta(k)$	the association event continuous random process vector representing the uncertainty in the origin of measurements
Θ_{i_k}	the association event discrete random process vector representing the uncertainty in the origin of measurements
$\{\Theta_{i_\ell}\}_1^{k-1}$	the history of association events from sample $\ell = 1$ through sample $\ell = k - 1$
$N_m(k)$	the number of measurements at sample k
N_{DT, i_k}	the number of measurements hypothesized under association event Θ_{i_k} as originating from targets hypothesized in a previous scan and detected in the current scan k
N_{TGT}	the total number of existing targets hypothesized under the association event history through sample $k - 1$
N_{NT, i_k}	the number of measurements hypothesized under association event Θ_{i_k} as originating from potential new targets at the current scan k
N_{FT, i_k}	the number of measurements hypothesized under association event Θ_{i_k} as originating from false sources at the current scan k
$N_H(k)$	the original number of hypothesized mixture components in the target state multivariate Gaussian mixture pdf at sample k before mixture reduction is applied
$N_R(k)$	the reduced number of mixture components in the target state multivariate Gaussian mixture pdf at sample k after mixture reduction is applied
$T\{pdf_1, pdf_2\}$	a true distance measure between two pdfs
$D\{pdf_1, pdf_2\}$	a pseudo-distance measure between two pdfs
$ \mathbf{A} = \det A$	the determinant of the matrix \mathbf{A}
$\langle f(x), g(x) \rangle$	the inner product of two functions of x defined as $\int_{x \in \mathbb{X}} f(x)g(x)dx$

GAUSSIAN MIXTURE REDUCTION FOR BAYESIAN TARGET TRACKING IN CLUTTER

I. Introduction

Measurement origin and target kinematics model parameter uncertainties are two fundamental problems encountered when estimating the state (position, velocity, etc.) of targets in clutter. By modeling these sources of uncertainty as random quantities, Bayes estimation may be used to estimate the unknown target state. If the uncertainty is modeled as a discrete random vector (i.e., unknown but constant over time), then the resulting Bayesian solution is computationally tractable. However, if the uncertainty is modeled as a discrete random process vector (i.e., unknown and generally changing over time), then the rigorous Bayesian solution is a summation of weighted Gaussian probability density functions in which the number of terms in the summation increases exponentially with time (assuming linear dynamics and measurement models, and normally distributed noise disturbances and initial state) [4,21,38]. Thus, this solution is computationally intractable, and approximation is necessary to implement the Bayesian solution.

Two types of Gaussian mixture approximations to the rigorous Bayesian solution can be used for reducing the number of mixture components (summation terms in the Gaussian mixture (pdf)) when the measurement origin uncertainty is modeled as a discrete random process vector. The first type of approximation is to reduce the Gaussian mixture pdf to a *single* Gaussian pdf at the end of each tracking system scan. This approximation is used in the Probabilistic Data Association Filter (PDAF) and the Joint Probabilistic Data Association Filter (JPDAF) [5]. Although this approximation is relatively simple and it does not require many computations, it is a rather crude approximation in cases in which the target state pdf is multi-modal with well-spaced peaks [31]. The other type of approximation is to reduce the number of

mixture components in the target state pdf at the end of each scan, as by using some type of measure function (providing a measure of the difference between the original pdf and the pdf with the reduced number of mixture components) to guide reduction decisions. This kind of approximation typically provides a better reduced-component representation of the original Gaussian mixture target state pdf, but at the expense of increased computational complexity and time.

Various examples of reducing the number of mixture components are found in the literature. Singer et al. [32] proposed merging measurement histories after $(N+1)$ scans and/or limiting the pool of measurements to include in the measurement association hypothesis generation process to decrease the number of mixture components. Reid [27] manages the explosive growth of mixture components by merging similar ones and deleting unlikely ones. Alspach [1], Lainiotis and Park [18], and Salmond [30] appear to be the first researchers to attack this problem by reducing the number of mixture components based on a measure function. The latter of this group proposed optimally reducing the number of components by minimizing a cost function based on the covariance matrix of the Gaussian mixture. Fourteen years later, Williams [38]¹ also proposed using a cost function criterion for mixture reduction. However, Williams' cost function differed from that of Salmond by considering the impact of reduction actions on the *entire* Gaussian mixture pdf, and not just the covariance matrix. Compared to other possible cost functions (such as the Kolmogorov Variational Distance [1, 2, 24, 28, 38], Bhattacharyya coefficient [18], and Kullback-Leibler distances [17]), Williams' cost function can be evaluated without approximation, yielding tractability. In fact, Williams' Integral Square Error (ISE)

¹This author owes a great deal of gratitude to Williams since the majority of this thesis is based on his work.

mixture reduction algorithm (MRA) produced what is thought to be the best single target tracking performance in a heavy clutter environment [38, 40, 41].

1.1 *Research Goal*

Recommendations in [38] indicate the potential for improving upon the results produced by Williams' ISE MRA. The goal of this research is to create a new MRA which offers better tracking performance and/or decreased computation time as compared to Williams' algorithm. Research conducted for this thesis will be tailored to meet this goal. Specifically, the fields of mathematical statistics and statistical inference will be explored to cultivate ideas which may be useful for developing the new MRA.

1.2 *Organization*

With the above goal in mind, the remainder of this thesis is organized into six additional chapters. Chapters II, III, and IV provide the background concepts needed for this thesis. The Bayesian approach to target tracking is covered in Chapter II, which introduces the concept of target tracking, target tracking models, target state pdf estimation, Gaussian mixture pdfs, and target state pdf estimation in the presence of kinematics model parameter uncertainty and measurement origin uncertainty. Chapter III is motivated by the recommendation in [38] to utilize a maximum likelihood estimation measure function. It summarizes the maximum likelihood estimation and Expectation Maximization techniques for estimating pdfs, and it also develops a maximum likelihood estimation-based measure function. The final background chapter, Chapter IV, presents distance and pseudo-distance measure functions for approximating pdfs as well as the MRAs developed by Williams [38, 40, 41] and Salmond [30, 31].

Based on the material covered in the background chapters, four new MRAs are developed, implemented, tested, and analyzed in Chapter V. The best performing MRA is chosen and tested in a single-target in heavy clutter tracking simulation

scenario presented in Chapter VI. Finally, Chapter VII concludes this thesis by summarizing the results of this research and recommending ideas for future study.

II. The Bayesian Approach to Target Tracking

Target tracking is a means of determining the state of moving objects over some time interval of interest from observations of the objects in the presence of uncertainty. A state may be one or several random processes that completely describes the behavior of the moving objects at any point in the time interval of interest. The movement of the objects is described by one or more *dynamics models* that mathematically characterize the motion of the objects. *Observations* of the objects are made using sensors. *Uncertainty* is present in both the objects' movement and in observations of the objects. Both sensor *measurement model* adequacy concerns and actual measurement corruption noise contribute to this uncertainty.

One common approach to target tracking is Bayesian estimation. This estimation method is used to obtain, in most instances, a real-time, recursive solution for a target tracking problem, which makes Bayesian estimation an ideal tool for target tracking. The goal of this chapter is to present the Bayesian approach to target tracking and highlight its versatility in solving problems which exhibit uncertainty in the target state, the dynamics model parameters, and the origin of measurements. Bayesian estimation uses Bayes' rule to solve for the target state pdf, and it will be shown that the Bayesian approach to target tracking may be boiled down to *one equation*. This insight has been mentioned in other sources [4, 5, 38], and Bayes' rule is emphasized as the starting point for solving every problem in this chapter.

Target tracking scenarios can be roughly categorized by the number of targets, the number of sensors, and the number of dynamics models. Scenarios which allow multiple targets are more prevalent in practice, whereas single target tracking scenarios are more often used as an instructional tool because of their relative simplicity. When available, multiple sensors usually provide more information about the state of the targets because the sensors' observations may be optimally combined using a multi-sensor fusion algorithm. In contrast, a single-sensor tracking system typically provides less information than a multi-sensor system using sensors with the same accuracy, but the tracking algorithms become less complicated than in a multi-sensor

system. In some cases, tracking system designers do not know the correct dynamics model to incorporate in their system, so multiple possible models are included in the design to account for this source of uncertainty. Consequently, the design architecture requires some type of multiple model estimation algorithm. In other instances, though, one dynamics model may be enough to handle a given tracking scenario.

Figure 2.1 depicts a block diagram of the target tracking process in which a presumed number of targets exist in the observation environment. At each scan, the sensors (or single sensor) obtain noise-corrupted measurements which are mathematically related to the state through the measurement model. The origin of each measurement is not known, and each measurement could belong to any one of the targets or it could have resulted from an erroneous detection due to clutter or sensor error (as characterized by its false alarm rate). Therefore, a tracking system may need to perform a measurement association process to reconcile the origin of each measurement to remove some uncertainty in the overall tracking problem. The measurements are fed into a bank of estimators (or one estimator if only one model is used), each with its own presumed dynamics model and hypothesized measurement association set, and a final state estimate is made by appropriately combining the outputs of the separate state estimators.

This chapter is organized as follows. Target kinematics models are presented in Section 2.1. Section 2.2 introduces recursive Bayesian filtering for linear and nonlinear estimation of the target state pdf. A description of the multivariate Gaussian mixture pdf is provided in Section 2.3 to lay the groundwork for understanding the Bayesian solution when uncertainty exists in the kinematics model parameters, Section 2.4, or in the source of the measurements, Section 2.5. Understanding Gaussian mixture pdfs will also be useful for Chapter III, Estimating Probability Density Functions, and Chapter IV, Approximating Gaussian Mixtures and Mixture Reduction.

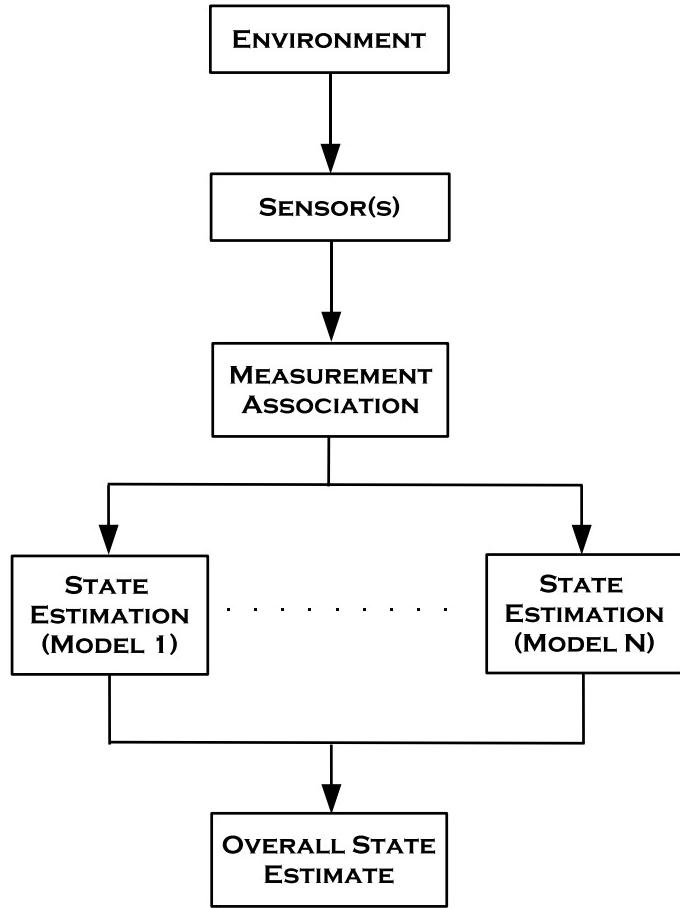


Figure 2.1: A conceptual block diagram of a target tracking algorithm.

2.1 Target Kinematics Models

Adequate target kinematics models are essential for accurate target tracking. Without sufficient models, even the most advanced model-based target tracking algorithm falls apart. Kinematics models, or dynamics models, are based on differential equations which characterize the evolution of the objects' movements over time. By its nature, a differential equation is only an approximation of the actual target dynamics. Since the mathematical descriptions are approximations, there is uncertainty in the fidelity of the model and the differential equations become stochastic in nature [21]. If the stochastic differential equations are restricted to the time-invariant class of problems, then these models may be succinctly written in state space form as (assuming no deterministic input)

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t) + \mathbf{G}\mathbf{w}(t) \quad (2.1)$$

where $\dot{\mathbf{x}}(t)$ is the time-derivative of the n -dimensional *state random process vector*¹, \mathbf{F} is the n -by- n time-invariant *system dynamics matrix*, \mathbf{G} is the n -by- s time-invariant *noise input matrix*, $\mathbf{w}(t)$ is the s -dimensional *model noise process vector* (assumed zero-mean, uncorrelated in time or “white,” and Gaussian) [21]. Equation (2.1) is called the *system dynamics* or *kinematics model*. Likewise, uncertain observations are made at time samples t_k and are given by the mathematical model

$$\mathbf{z}(t_k) = \mathbf{H}\mathbf{x}(t_k) + \mathbf{v}(t_k) \quad (2.2)$$

where $\mathbf{z}(t_k)$ is the m -dimensional *measurement random process vector* at sample time t_k , \mathbf{H} is the m -by- n time-invariant *measurement matrix*, and $\mathbf{v}(t_k)$ is the m -dimensional *measurement noise process vector* (also assumed zero-mean, white, and Gaussian) [21]. Equation (2.2) is termed the *measurement model*. A set of initial conditions must be specified to obtain a particular solution to the above differential

¹The state random process vector may represent kinematic quantities such as position, velocity, and acceleration.

equations. This set of initial conditions is usually unknown, in which case it may be represented by a Gaussian random vector with mean and covariance specified by

$$\begin{aligned} E\{\mathbf{x}(t_0)\} &= \hat{\mathbf{x}}_0 \\ E\{[\mathbf{x}(t_0) - \hat{\mathbf{x}}_0][\mathbf{x}(t_0) - \hat{\mathbf{x}}_0]^T\} &= \mathbf{P}_0. \end{aligned} \quad (2.3)$$

Eventually, these models will likely be implemented on a computer, and a discrete-time form for Equation (2.1) will be required. If k is the sample index², then the shift-invariant³ discrete-time models are

$$\begin{aligned} \text{System Dynamics Model: } \mathbf{x}(k) &= \Phi(k, k-1)\mathbf{x}(k-1) + \mathbf{G}_d\mathbf{w}_d(k-1) \\ \text{Measurement Model: } \mathbf{z}(k) &= \mathbf{H}\mathbf{x}(k) + \mathbf{v}(k) \end{aligned} \quad (2.4)$$

where the previous nomenclature holds, d stands for “discrete-time,” and $\Phi(k, k-1)$ is the n -by- n discrete-time *state transition matrix* given by $\Phi(k, k-1) = e^{\mathbf{F}(t_k - t_{k-1})}$ [21]. The initial conditions to these difference equations are

$$\begin{aligned} E\{\mathbf{x}(0)\} &= \hat{\mathbf{x}}_0 \\ E\{[\mathbf{x}(0) - \hat{\mathbf{x}}_0][\mathbf{x}(0) - \hat{\mathbf{x}}_0]^T\} &= \mathbf{P}_0. \end{aligned} \quad (2.5)$$

²The sampling interval is defined as $T = t_k - t_{k-1}$.

³In the discrete-time formulation, “time-invariant” becomes “shift-invariant.”

The first- and second-order statistics of the noise process vectors are

$$\begin{aligned}
E\{\mathbf{w}_d(k)\} &= \mathbf{0} \\
E\{\mathbf{w}_d(k)\mathbf{w}_d^T(l)\} &= \mathbf{Q}_d(k) \delta_{kl} \\
E\{\mathbf{v}(k)\} &= \mathbf{0} \\
E\{\mathbf{v}(k)\mathbf{v}^T(l)\} &= \mathbf{R}(k) \delta_{kl} \\
E\{\mathbf{v}(k)\mathbf{w}_d^T(l)\} &= \mathbf{0} \\
E\{\mathbf{v}(k)\mathbf{x}^T(0)\} &= \mathbf{0} \\
E\{\mathbf{w}_d(l)\mathbf{x}^T(0)\} &= \mathbf{0}
\end{aligned} \tag{2.6}$$

where δ_{kl} is the Kronecker delta function:

$$\delta_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise.} \end{cases}$$

Since $\mathbf{x}(0)$, $\mathbf{w}_d(k)$, and $\mathbf{v}(k)$ are assumed jointly Gaussian⁴, the last three lines of (2.6) imply that they are independent.

Before continuing to the next section, it should be noted that nonlinear kinematics models are also possible. In discrete time, the nonlinear shift-invariant models are [13, 21]

$$\begin{aligned}
\text{System Dynamics Model: } \mathbf{x}(k) &= \boldsymbol{\phi}[\mathbf{x}(k-1)] + \mathbf{G}_d \mathbf{w}_d(k-1) \\
\text{Measurement Model: } \mathbf{z}(k) &= \mathbf{h}[\mathbf{x}(k)] + \mathbf{v}(k)
\end{aligned} \tag{2.7}$$

where $\boldsymbol{\phi}[\cdot]$ is the *nonlinear system dynamics vector function* and $\mathbf{h}[\cdot]$ is the *nonlinear measurement vector function*. Note that, in general, these equations may be shift-varying, in which case $\boldsymbol{\phi}[\cdot]$ and $\mathbf{h}[\cdot]$ would also be functions of the appropriate time index.

⁴If two jointly Gaussian random vectors are uncorrelated, then they are also independent. In general, though, uncorrelated random vectors are not necessarily independent [20, 21, 35].

One common linear system dynamics model used in target tracking is the *constant velocity* (CV) model [7, 21]⁵. The CV model is of particular interest since it was used in simulations by Salmond [30, 31] and Williams [38, 40, 41] to evaluate their Gaussian mixture reduction methods and it will be used in simulations to evaluate new Gaussian mixture reduction approaches later in this thesis. For a target traveling in the x - y plane, the *discrete-time* CV model is

$$\begin{bmatrix} x(k) \\ v_x(k) \\ y(k) \\ v_y(k) \end{bmatrix} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ v_x(k-1) \\ y(k-1) \\ v_y(k-1) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{bmatrix} \begin{bmatrix} w_x(k-1) \\ w_y(k-1) \end{bmatrix} \quad (2.8)$$

where T is the sampling interval, $x(k)$ and $y(k)$ are the x and y target positions, $v_x(k)$ and $v_y(k)$ are the target velocities in the x and y directions, and $w_x(k-1)$ and $w_y(k-1)$ are the model noise processes in the x and y coordinates. The model noise process is zero mean, and its covariance matrix is

$$\mathbf{Q}_d(k) = \mathbf{Q}_d = \begin{bmatrix} q_{d_x} & 0 \\ 0 & q_{d_y} \end{bmatrix}. \quad (2.9)$$

⁵Various other system dynamics models, as well as measurement models, may be found in [4, 5, 6, 7, 23, 33].

2.2 Recursive Bayesian Filtering

A recursive Bayesian filter⁶ is used to calculate the target state $\mathbf{x}(k)$. For linear models, the linear recursive Bayesian filter, which is the well-known Kalman filter, provides an optimal target state estimate by almost all practical criteria [21]. This filter is the topic of Subsection 2.2.1. Nonlinear recursive Bayesian filters are used when the dynamics models are nonlinear, the measurement models are nonlinear, or both, but in general nonlinear filters do not enjoy the same claim to optimality as the Kalman filter and require approximations to yield a finite-dimensional form. Nonlinear recursive Bayesian filters are the subject of Subsection 2.2.2.

2.2.1 Linear Recursive Bayesian Filtering. For the discrete-time linear models listed in Section 2.1, the Kalman filter is the optimal estimator of the state random process vector [21]. The Kalman filter may be derived from a Bayesian point of view, as in [21], or from an orthogonal projection perspective, as in [13, 35]. This subsection provides the key steps from [21] in the derivation of the Kalman filter using the Bayes estimation technique. These derivation steps will hopefully provide the reader with insights into the Kalman filter equations.

Before beginning the derivation, it is worthwhile to introduce important preliminary information about the Kalman filter. First, each iteration of the Kalman filter operates in two stages that are referred to as *time propagation* and *measurement update*. The notation “(estimate at time index|measurement history through time index)” allows one to follow the time varying quantities of the filter (e.g., $\hat{\mathbf{x}}(\cdot|\cdot)$, which is the state random process vector mean estimate) at each stage of a filter iteration. For example, the notation $\hat{\mathbf{x}}(k|k - 1)$ signifies that the state mean estimate has been propagated in time through the kinematics model to sample k , but measurements at sample k have not yet been incorporated into the estimate. Second, a composite vector \mathbf{Z}^k , called the *measurement history*, is defined as a composite vector formed from all

⁶The word “filter,” as it is used in the context of estimation, differs from the same word used in deterministic signal processing. In this thesis, a filter should be thought of as a type of estimator that includes observations up to and including sample k to determine the target state [13].

of the measurement random process vectors from the time of the initial measurement through the current sample, k [21]. That is,

$$\mathbf{Z}^k = \begin{bmatrix} \mathbf{z}(1) \\ \vdots \\ \mathbf{z}(k) \end{bmatrix}. \quad (2.10)$$

Likewise, a composite vector of *realized* measurements \mathbf{Z}^k , termed the *realized measurement history*, is defined as the vector containing all of the actual measurements through sample k , and is given by

$$\mathbf{Z}^k = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \end{bmatrix} \quad (2.11)$$

where \mathbf{z}_i , $i = 1, \dots, k$, is the i^{th} realized measurement random process vector [21]. Third, the initial state random process vector $\mathbf{x}(0)$ is assumed to be Gaussian, and since the dynamics and measurement models are linear, the state random process vector at any sample k is also Gaussian⁷. Additionally, a multivariate Gaussian pdf is completely specified by the mean and covariance parameters of the Gaussian random vector it represents, and expressions for these parameters can be found by applying the expectation operation, $E\{\cdot\}$, to the discrete-time models given in Equation (2.4) at the appropriate time index. Thus, the pdf of the state random process vector $\mathbf{x}(k)$ may be found at any sample k .

Using this preliminary information, the derivation of the Kalman filter equations begins by identifying the conditional pdf of the Gaussian state random process vector

⁷Linear transformations (e.g., the linear dynamics and measurement models) of a Gaussian random vector are also Gaussian [21], and linear combinations of *jointly* Gaussian random vectors are similarly Gaussian.

$\mathbf{x}(k-1)$ conditioned on the measurement history through sample $k-1$ as [21]

$$f(\mathbf{x}(k-1)|\mathbf{Z}^{k-1}) = \frac{\exp\left[-\frac{1}{2}(\cdot)^T \mathbf{P}(k-1|k-1)^{-1}(\cdot)\right]}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{P}(k-1|k-1)}} \quad (2.12)$$

$$(\cdot) = \mathbf{x}(k-1) - \hat{\mathbf{x}}(k-1|k-1)$$

or

$$f(\mathbf{x}(k-1)|\mathbf{Z}^{k-1}) \triangleq \mathcal{N}\{\mathbf{x}(k-1); \hat{\mathbf{x}}(k-1|k-1), \mathbf{P}(k-1|k-1)\}. \quad (2.13)$$

The values of $\hat{\mathbf{x}}(k-1|k-1)$ and $\mathbf{P}(k-1|k-1)$ are the outputs from the last iteration of the Kalman filter.

Next, the state random process vector $\mathbf{x}(k-1)$ is propagated in time through the linear dynamics model (the first line of Equation (2.4)) in the *time propagation* stage. Since the model is linear and the state and model noise process vectors are jointly Gaussian, the state random process vector at sample k conditioned on the measurement history through sample $k-1$ is also Gaussian [21]. This fact can be shown by applying Bayes' rule⁸ and the result that the product of two Gaussian pdfs is a Gaussian pdf [21]. The state random process vector pdf is given by $\mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}(k|k-1), \mathbf{P}(k|k-1)\}$, and its mean and covariance parameters are

$$\hat{\mathbf{x}}(k|k-1) = \Phi(k, k-1)\hat{\mathbf{x}}(k-1|k-1) \quad (2.14)$$

$$\mathbf{P}(k|k-1) = \Phi(k, k-1)\mathbf{P}(k-1|k-1)\Phi(k, k-1)^T + \mathbf{G}_d\mathbf{Q}_d(k-1)\mathbf{G}_d^T.$$

⁸The law of conditional probability for two pdfs states that

$$f(A|B) = \frac{f(AB)}{f(B)}.$$

Bayes' rule follows from the law of conditional probability for pdfs and is given by

$$f(A|B) = \frac{f(B|A)f(A)}{f(B)}.$$

Equations (2.14) are the *Kalman filter time propagation equations*.

The next step in the Kalman filter equations derivation is to apply the current measurement, $\mathbf{z}(k)$, in the *measurement update* stage. Using Bayes' rule and noting that Equation (2.10) may be written as

$$\begin{bmatrix} \mathbf{Z}^{k-1} \\ \mathbf{z}(k) \end{bmatrix}$$

the pdf of the state random process vector conditioned on *all* of the measurements through sample k is [21]

$$\begin{aligned} f(\mathbf{x}(k)|\mathbf{Z}^k) &= \frac{f(\mathbf{x}(k), \mathbf{z}(k), \mathbf{Z}^{k-1})}{f(\mathbf{z}(k), \mathbf{Z}^{k-1})} \quad (\text{conditional probability for pdfs}) \\ &= \frac{f(\mathbf{z}(k)|\mathbf{x}(k), \mathbf{Z}^{k-1}) f(\mathbf{x}(k), \mathbf{Z}^{k-1})}{f(\mathbf{z}(k)|\mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})} \quad (\text{more of the same}) \\ &= \frac{f(\mathbf{z}(k)|\mathbf{x}(k), \mathbf{Z}^{k-1}) f(\mathbf{x}(k)|\mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})}{f(\mathbf{z}(k)|\mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})} \quad (\dots \text{and again}) \\ &= \frac{f(\mathbf{z}(k)|\mathbf{x}(k), \mathbf{Z}^{k-1}) \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}(k|k-1), \mathbf{P}(k|k-1)\}}{f(\mathbf{z}(k)|\mathbf{Z}^{k-1})}. \end{aligned} \quad (2.15)$$

From a broad perspective, the last line of Equation (2.15) provides the justification for the term “*recursive Bayesian filter*.” The state random process vector pdf at the current sample k , conditioned on the latest measurements, cannot be calculated until the state random process vector pdf conditioned on the measurements up to the *previous* time index is determined. That is, Equation (2.14) must first be determined before the pdf $\mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}(k|k-1), \mathbf{P}(k|k-1)\}$ can be specified. Hence, the recursive nature of the filter is apparent in the last line of Equation (2.15). The terms $f(\mathbf{z}(k)|\mathbf{x}(k), \mathbf{Z}^{k-1})$ and $f(\mathbf{z}(k)|\mathbf{Z}^{k-1})$ are found with relative ease by noting that they are Gaussian pdfs that are completely specified by their respective mean and covariance [21]. One may find these parameters by applying the conditional ex-

pectation operation to the discrete-time measurement model given in the second line of Equation (2.4). However, determining the final form of $f(\mathbf{x}(k)|\mathbf{Z}^k)$ is not so easy since it requires a rather unpleasant amount of linear algebraic manipulations to show that $f(\mathbf{x}(k)|\mathbf{Z}^k)$ is a Gaussian pdf.

Returning to the derivation, the initial representation of Equation (2.15) is a complicated algebraic form which has very little apparent resemblance to a Gaussian pdf [21]. However, after several pages of algebraic manipulations, $f(\mathbf{x}(k)|\mathbf{Z}^k)$ assumes the much nicer form of [21]

$$\begin{aligned} f(\mathbf{x}(k)|\mathbf{Z}^k) &= \frac{\exp\left[-\frac{1}{2}[\mathbf{x}(k) - \hat{\mathbf{x}}(k|k)]^T \mathbf{P}(k|k)^{-1} [\mathbf{x}(k) - \hat{\mathbf{x}}(k|k)]\right]}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{P}(k|k)}} \\ &= \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}(k|k), \mathbf{P}(k|k)\}. \end{aligned}$$

The parameters of this pdf are (note that \mathbf{z}_k is the realization of the random vector $\mathbf{z}(k)$ at sample k)

$$\begin{aligned} \hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}(k|k-1)] \\ \mathbf{P}(k|k) &= \mathbf{P}(k|k-1) - \mathbf{K}(k)\mathbf{H}\mathbf{P}(k|k-1) \end{aligned} \quad (2.16)$$

and they are referred to as the *Kalman filter measurement update equations* [21]. Note that the term $\mathbf{H}\hat{\mathbf{x}}(k|k-1) = \hat{\mathbf{z}}(k|k-1)$ is the conditional mean of the measurement at sample k , or the *predicted measurement*. Also, the term $\mathbf{K}(k)$ is called the *Kalman gain* which is given by [21]

$$\mathbf{K}(k) = \mathbf{P}(k|k-1)\mathbf{H}^T[\mathbf{H}\mathbf{P}(k|k-1)\mathbf{H}^T + \mathbf{R}(k)]^{-1}. \quad (2.17)$$

The previous equation completes the presentation of the key steps in the derivation of the Kalman filter equations using the Bayes estimation technique.

In summary, given the discrete-time models in Equation (2.4), including the initial conditions in Equation (2.5) and noise statistics given in Equation (2.6), one iteration of the Kalman filter is calculated by using the *Kalman filter time propagation equations*:

- Propagate the mean estimate to sample k according to the system dynamics model:

$$\hat{\mathbf{x}}(k|k-1) = \Phi(k, k-1)\hat{\mathbf{x}}(k-1|k-1)$$

- Propagate the covariance estimate to sample k according to the system dynamics model and add the covariance of the model noise process vector:

$$\mathbf{P}(k|k-1) = \Phi(k, k-1)\mathbf{P}(k-1|k-1)\Phi(k, k-1)^T + \mathbf{G}_d\mathbf{Q}_d(k-1)\mathbf{G}_d^T,$$

followed by the *Kalman filter measurement update equations*:

- Calculate the covariance of the *residual* $\mathbf{r}(k)$ and the realized *residual* \mathbf{r}_k , which is the residual evaluated with the observed measurement \mathbf{z}_k :

$$\begin{aligned} \mathbf{r}(k) &= \mathbf{z}(k) - \hat{\mathbf{z}}(k|k-1) \\ \mathbf{S}(k) &= E\{\mathbf{r}(k)\mathbf{r}^T(k)\} \\ &= \mathbf{H}\mathbf{P}(k|k-1)\mathbf{H}^T + \mathbf{R}(k) \\ \mathbf{r}(k)|_{\mathbf{z}(k)=\mathbf{z}_k} &= \mathbf{r}_k \\ &= \mathbf{z}_k - \hat{\mathbf{z}}(k|k-1) \end{aligned}$$

- Compute the Kalman gain as a function of the uncertainty in the system dynamics model and the measurement model at sample k :

$$\mathbf{K}(k) = \mathbf{P}(k|k-1)\mathbf{H}^T\mathbf{S}^{-1}(k)$$

- Add the “new information” from the measurement at sample k to the propagated mean estimate to form the measurement-updated conditional mean estimate:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\mathbf{r}_k$$

- Subtract the *weighted* propagated covariance estimate from the propagated covariance estimate:

$$\mathbf{P}(k|k) = \mathbf{P}(k|k-1) - \mathbf{K}(k)\mathbf{H}\mathbf{P}(k|k-1).$$

The Kalman gain plays a pivotal role in the Kalman filter because it reflects the amount of confidence placed in the information provided by the measurements relative to that of the propagated information [35]. If the uncertainty in the measurements is large (as indicated by $\mathbf{R}(k)$), then the Kalman gain is small (notice $[\dots + \mathbf{R}(k)]^{-1}$ in Equation (2.17)), and the impact of the measurement on the state random process vector mean and covariance estimates is small. Likewise, if the measurement uncertainty is small, then the entries in the matrix $\mathbf{R}(k)$ are small, and the measurements have a greater impact on the updated state random process vector mean and covariance estimates.

From a visual perspective, the conditional pdfs $f(\mathbf{x}(k-1)|\mathbf{Z}^{k-1})$, $f(\mathbf{x}(k)|\mathbf{Z}^{k-1})$, and $f(\mathbf{x}(k)|\mathbf{Z}^k)$ are modified according to the Kalman filter equations. For a typical problem involving a scalar state random process, these pdfs appear in Figure 2.2. Initially, before the time propagation and measurement update stages are entered, the scalar state process has a conditional pdf represented by the dotted trace in the figure. After time propagation, the scalar state process pdf, shown as the dash-dotted trace, is modified according to Equation (2.14) and the width of the pdf is larger than before since the dynamics model adds uncertainty. Finally, the measurement is incorporated and the scalar state random process pdf is narrower than the propagated pdf as shown by the solid trace. This narrowing of the pdf is expected since the

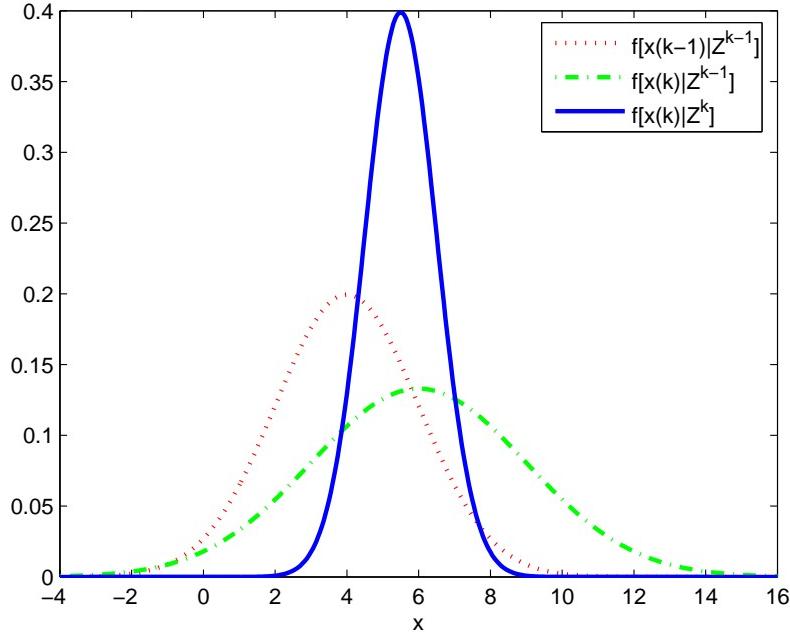


Figure 2.2: Typical conditional scalar state pdfs before time propagation, after time propagation, and after measurement update.

variance of the scalar state random process is smaller than that of the propagated scalar state random process according to Equation (2.16): this is the benefit obtained from the most recent measurement.

2.2.2 Nonlinear Recursive Bayesian Filtering. Nonlinear recursive Bayesian filters are used when the dynamics and/or measurement models are nonlinear. Nonlinear transformations destroy the Gaussian nature of the target state random process vector, and the mean and covariance of the state random process vector no longer completely describe the target state pdf. In the best case, if the nonlinearity in the transformation is negligible, then a Gaussian pdf may be a good approximation to the true target state pdf. In the worst case, if the nonlinearity in the transformation is substantial, then the true target state pdf will likely bear little resemblance to a Gaussian pdf. Considering either case, an *optimal* nonlinear recursive Bayesian filter would need to compute an infinite number of moments to characterize the *exact* target state pdf [16, 22]. By contrast, only the mean and covariance of the state ran-

dom process vector are required to describe the target state pdf completely when the models are linear.

Tracking problems that require *slightly* nonlinear models are adequately solved with the extended Kalman filter (EKF). Problems which utilize *moderately* nonlinear models, or models in which analytic expressions for the Jacobian of the nonlinear system dynamics or nonlinear measurement vector functions do not exist⁹, are potentially better suited to an unscented Kalman filter solution. In the case of *highly* nonlinear models, or models in which closed-form expressions for the Jacobian of the nonlinear system dynamics or nonlinear measurement vector functions are not available, the tracking problem could be solved using a particle filter. The EKF is introduced in this subsection, but the unscented Kalman filter and the particle filter are not covered.

For nonlinear models, both the nonlinear system dynamics vector function and nonlinear measurement vector function can be written in a Taylor series expansion about some nominal point as long as an analytic expression for the Jacobian and higher-order derivatives of both nonlinear vector functions exist. If the vector functions are slightly nonlinear, then the second- and higher-order terms in their respective expansions may be justifiably ignored to create a first-order *linear* approximation of each nonlinear vector function [22]. Then, the Kalman filter Equations (2.14), (2.16), and (2.17) can be used by replacing the state transition and measurement matrices with the Jacobian of the nonlinear system dynamics and measurement vectors, respectively, about some appropriate point [21]. These steps form the basis for the EKF derivation.

A mathematical derivation of the EKF similar to the one found in [22] is presented below. The final equations for the EKF appear at the end of this subsection, and they are placed in the same propagate-update structure as used for the Kalman

⁹In [14], the author proposes using radar cross-section measurements to track and identify targets simultaneously. Because of the nature of the measurement equation for radar cross-section, a closed-form measurement equation does not exist. Consequently, the Jacobian of the measurement equation cannot be found.

filter equations to accentuate the similarity between the steps required to implement each filter. These equations emphasize an important distinction between the Kalman filter and the EKF: unlike in the Kalman filter equations, the state mean estimate, $\hat{\mathbf{x}}(\cdot)$, and state covariance estimate, $\mathbf{P}(\cdot)$, are interdependent in EKF equations. The interdependence of the state mean and covariance estimates stems from expanding the Taylor series about the state mean estimate. Also, since the *true* models are nonlinear, one should keep in mind that the *true* target state pdf is not a Gaussian pdf as one may be led to believe by the EKF equations.

The nonlinear, shift-invariant system dynamics and measurement models are given as

$$\begin{aligned} \text{System Dynamics Model: } & \mathbf{x}(k) = \phi[\mathbf{x}(k-1)] + \mathbf{G}_d \mathbf{w}_d(k-1) \\ \text{Measurement Model: } & \mathbf{z}(k) = \mathbf{h}[\mathbf{x}(k)] + \mathbf{v}(k). \end{aligned} \quad (2.7)$$

If $\hat{\mathbf{x}}(k-1|k-1)$ is the conditional mean of $\mathbf{x}(k-1)$, and $\tilde{\mathbf{x}}(k-1)$ is described by the first- and second-order statistics

$$\begin{aligned} E\{\tilde{\mathbf{x}}(k-1)|\mathbf{Z}^{k-1}\} &= \mathbf{0} \\ E\{\tilde{\mathbf{x}}(k-1)\tilde{\mathbf{x}}(k-1)^T|\mathbf{Z}^{k-1}\} &= \mathbf{P}(k-1|k-1) \end{aligned}$$

then the state random process vector at sample $k-1$ is equivalently represented by $\mathbf{x}(k-1) = \hat{\mathbf{x}}(k-1|k-1) + \tilde{\mathbf{x}}(k-1)$. In a similar manner, the propagated state random process vector at sample k may be written as $\mathbf{x}(k) = \hat{\mathbf{x}}(k|k-1) + \tilde{\mathbf{x}}(k)$, where $\tilde{\mathbf{x}}(k)$ is zero-mean and $\mathbf{P}(k|k-1)$ is its covariance. The Taylor series expansion of

$\phi[\mathbf{x}(k-1)]$ about $\hat{\mathbf{x}}(k-1|k-1)$ is¹⁰

$$\phi[\mathbf{x}(k-1)] = \phi[\hat{\mathbf{x}}(k-1|k-1) + \tilde{\mathbf{x}}(k-1)]$$

$$\begin{aligned} &= \phi[\hat{\mathbf{x}}(k-1|k-1)] + \frac{\partial \phi[\mathbf{x}(k-1)]}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \tilde{\mathbf{x}}(k-1) + \text{H.O.T.} \\ &\approx \phi[\hat{\mathbf{x}}(k-1|k-1)] + \frac{\partial \phi[\mathbf{x}(k-1)]}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \tilde{\mathbf{x}}(k-1) \end{aligned} \quad (2.18)$$

where H.O.T. stands for “higher order terms” which are neglected in the first-order approximation [3, 22]. A similar expression may be found for the first-order linear approximation to the nonlinear measurement vector function.

The propagated state conditional mean and covariance estimates may be found by substituting Equation (2.18) into the conditional expectation equations for these quantities, conditioned on the measurement history through sample $k-1$. Noting that $\tilde{\mathbf{x}}(\cdot)$ is zero-mean, $\mathbf{w}_d(\cdot)$ is zero-mean and uncorrelated in time, and $\phi[\hat{\mathbf{x}}(k-1|k-1)]$

¹⁰A modified form of the Taylor series is given in [3] as

$$\phi(x+h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} \phi^{(n)}(x).$$

This form is adapted according to [22] and used in the derivation of the propagation equations for the EKF.

is a constant, the propagated state mean estimate is

$$\begin{aligned}
\hat{\mathbf{x}}(k|k-1) &= E \{ \mathbf{x}(k) | \mathbf{Z}^{k-1} \} \\
&= E \{ \phi[\mathbf{x}(k-1)] + \mathbf{G}_d \mathbf{w}_d(k-1) | \mathbf{Z}^{k-1} \} \\
&\approx \phi[\hat{\mathbf{x}}(k-1|k-1)] + \mathbf{G}_d E \{ \mathbf{w}_d(k-1) | \mathbf{Z}^{k-1} \} \\
&\quad + \frac{\partial \phi[\mathbf{x}(k-1)]}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} E \left\{ \tilde{\mathbf{x}}(k-1) \Big| \mathbf{Z}^{k-1} \right\} \\
&= \phi[\hat{\mathbf{x}}(k-1|k-1)]. \tag{2.19}
\end{aligned}$$

In a similar manner, the propagated state random process vector covariance estimate is

$$\begin{aligned}
\mathbf{P}(k|k-1) &= E \{ [\mathbf{x}(k) - \hat{\mathbf{x}}(k|k-1)][\mathbf{x}(k) - \hat{\mathbf{x}}(k|k-1)]^T | \mathbf{Z}^{k-1} \} \\
&= E \{ \mathbf{x}(k) \mathbf{x}(k)^T | \mathbf{Z}^{k-1} \} - \hat{\mathbf{x}}(k|k-1) \hat{\mathbf{x}}(k|k-1)^T \\
&= E \{ [\phi[\mathbf{x}(k-1)] + \mathbf{G}_d \mathbf{w}_d(k-1)][\phi[\mathbf{x}(k-1)] + \mathbf{G}_d \mathbf{w}_d(k-1)]^T | \mathbf{Z}^{k-1} \} \\
&\quad - \hat{\mathbf{x}}(k|k-1) \hat{\mathbf{x}}(k|k-1)^T \\
&= E \{ \phi[\mathbf{x}(k-1)] \phi[\mathbf{x}(k-1)]^T | \mathbf{Z}^{k-1} \} \\
&\quad + E \{ \phi[\mathbf{x}(k-1)] \mathbf{w}_d(k-1)^T \mathbf{G}_d^T | \mathbf{Z}^{k-1} \} \\
&\quad + E \{ \mathbf{G}_d \mathbf{w}_d(k-1) \phi[\mathbf{x}(k-1)]^T | \mathbf{Z}^{k-1} \} \\
&\quad + \mathbf{G}_d E \{ \mathbf{w}_d(k-1) \mathbf{w}_d(k-1)^T | \mathbf{Z}^{k-1} \} \mathbf{G}_d^T - \hat{\mathbf{x}}(k|k-1) \hat{\mathbf{x}}(k|k-1)^T
\end{aligned}$$

Aside 1:

$$\begin{aligned}
E\{\phi[\mathbf{x}(k-1)]\phi[\mathbf{x}(k-1)]^T|\mathbf{Z}^{k-1}\} &\approx E\{\phi[\hat{\mathbf{x}}(k-1|k-1)]\phi[\hat{\mathbf{x}}(k-1|k-1)]^T|\mathbf{Z}^{k-1}\} \\
&+ E\left\{\phi[\hat{\mathbf{x}}(k-1|k-1)]\tilde{\mathbf{x}}(k-1)^T \frac{\partial\phi[\mathbf{x}(k-1)]^T}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \Big| \mathbf{Z}^{k-1}\right\} \\
&+ E\left\{\frac{\partial\phi[\mathbf{x}(k-1)]}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \tilde{\mathbf{x}}(k-1)\phi[\hat{\mathbf{x}}(k-1|k-1)]^T \Big| \mathbf{Z}^{k-1}\right\} \\
&+ \frac{\partial\phi[\mathbf{x}(k-1)]}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \cdot E\{\tilde{\mathbf{x}}(k-1)\tilde{\mathbf{x}}(k-1)^T|\mathbf{Z}^{k-1}\} \\
&\cdot \frac{\partial\phi[\mathbf{x}(k-1)]^T}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \\
&= \phi[\hat{\mathbf{x}}(k-1|k-1)]\phi[\hat{\mathbf{x}}(k-1|k-1)]^T + \mathbf{0} + \mathbf{0} + \\
&\frac{\partial\phi[\mathbf{x}(k-1)]}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \cdot \mathbf{P}(k-1|k-1) \cdot \frac{\partial\phi[\mathbf{x}(k-1)]^T}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \\
&= \hat{\mathbf{x}}(k|k-1)\hat{\mathbf{x}}(k|k-1)^T + \\
&\frac{\partial\phi[\mathbf{x}(k-1)]}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \cdot \mathbf{P}(k-1|k-1) \cdot \frac{\partial\phi[\mathbf{x}(k-1)]^T}{\partial\mathbf{x}(k-1)} \Big|_{x(k-1)=\hat{\mathbf{x}}(k-1|k-1)}
\end{aligned}$$

Aside 2:

$$\begin{aligned}
E\{\phi[\mathbf{x}(k-1)]\mathbf{w}_d(k-1)^T\mathbf{G}_d^T|\mathbf{Z}^{k-1}\} &= E\{\phi[\mathbf{x}(k-1)]|\mathbf{Z}^{k-1}\} E\{\mathbf{w}_d(k-1)^T|\mathbf{Z}^{k-1}\} \mathbf{G}_d^T \\
&= \mathbf{0} \\
E\{\mathbf{G}_d\mathbf{w}_d(k-1)\phi[\mathbf{x}(k-1)]^T|\mathbf{Z}^{k-1}\} &= \mathbf{0}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(k|k-1) &= \frac{\partial \phi[\mathbf{x}(k-1)]}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \cdot \mathbf{P}(k-1|k-1) \\
&\cdot \frac{\partial \phi[\mathbf{x}(k-1)]^T}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \\
&+ \mathbf{G}_d \mathbf{Q}_d \mathbf{G}_d^T + \hat{\mathbf{x}}(k|k-1) \hat{\mathbf{x}}(k|k-1)^T - \hat{\mathbf{x}}(k|k-1) \hat{\mathbf{x}}(k|k-1)^T \\
&= \frac{\partial \phi[\mathbf{x}(k-1)]}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \cdot \mathbf{P}(k-1|k-1) \cdot \frac{\partial \phi[\mathbf{x}(k-1)]^T}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \\
&+ \mathbf{G}_d \mathbf{Q}_d \mathbf{G}_d^T. \tag{2.20}
\end{aligned}$$

Expressions for the measurement-updated state mean and covariance estimates may be derived using the same techniques used in the propagated estimates derivation.

The final EKF equations are listed below according to the EKF stage in which they are calculated. At the *time propagation stage* of the EKF, calculate

$$\begin{aligned}
\Phi(k, k-1) &= \frac{\partial \phi[\mathbf{x}(k-1)]}{\partial \mathbf{x}(k-1)} \Big|_{\mathbf{x}(k-1)=\hat{\mathbf{x}}(k-1|k-1)} \\
\hat{\mathbf{x}}(k|k-1) &= \phi[\hat{\mathbf{x}}(k-1|k-1)] \\
\mathbf{P}(k|k-1) &= \Phi(k, k-1) \mathbf{P}(k-1|k-1) \Phi(k, k-1)^T + \mathbf{G}_d \mathbf{Q}_d(k-1) \mathbf{G}_d^T. \tag{2.21}
\end{aligned}$$

In the *measurement update stage* of the EKF, calculate

$$\begin{aligned}
\mathbf{H}(k) &= \frac{\partial \mathbf{h}[\mathbf{x}(k)]}{\partial \mathbf{x}(k)} \Big|_{\mathbf{x}(k)=\hat{\mathbf{x}}(k|k-1)} \\
\mathbf{K}(k) &= \mathbf{P}(k|k-1) \mathbf{H}(k)^T [\mathbf{H}(k) \mathbf{P}(k|k-1) \mathbf{H}(k)^T + \mathbf{R}(k)]^{-1} \\
\hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k) [\mathbf{z}_k - \mathbf{h}[\hat{\mathbf{x}}(k|k-1)]] \\
\mathbf{P}(k|k) &= \mathbf{P}(k|k-1) - \mathbf{K}(k) \mathbf{H}(k) \mathbf{P}(k|k-1). \tag{2.22}
\end{aligned}$$

2.3 Multivariate Gaussian Mixtures

Understanding multivariate Gaussian mixtures will aid in understanding the target state pdf generated by the Bayesian solution for tracking problems in which uncertainty exists in the model parameters or in the origin of the measurements. More importantly, since Gaussian mixture reduction is the focus of this thesis, a good introduction to this subject is necessary before continuing to the next chapters.

In this thesis, a multivariate Gaussian mixture pdf is a weighted, finite sum of multivariate Gaussian pdfs. It is characterized by the number of mixture components and the weight, mean vector, and covariance matrix of each component. Since a pdf must be nonnegative, and the integral of a pdf over the sample space of the random quantity it represents must evaluate to unity, the mixture weights must be nonnegative and the sum of all of the weights must equal one. The multivariate Gaussian mixture pdf of the random vector \mathbf{x} with the parameter set Ω is represented by

$$f(\mathbf{x}|\Omega) = \sum_{i=1}^M p_i f(\mathbf{x}|\boldsymbol{\mu}_i, \mathbf{P}_i) \quad (2.23)$$

where M is the number of mixture components and p_i , $\boldsymbol{\mu}_i$, and \mathbf{P}_i are the weight, mean vector, and covariance matrix for each component $i = 1, \dots, M$. Each multivariate Gaussian pdf has the form

$$f(\mathbf{x}|\boldsymbol{\mu}_i, \mathbf{P}_i) = \mathcal{N}\{\mathbf{x}; \boldsymbol{\mu}_i, \mathbf{P}_i\} = \frac{\exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{P}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right]}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{P}_i}} \quad (2.24)$$

where n is the dimension of the random vector \mathbf{x} and the covariance \mathbf{P}_i is symmetric positive definite. Figure 2.3 illustrates a four-component univariate Gaussian mixture pdf.

The overall mean and covariance of \mathbf{x} given in (2.23) are derived in Appendix B of [31] and in Chapter II of [38]. These statistics are reproduced in the following

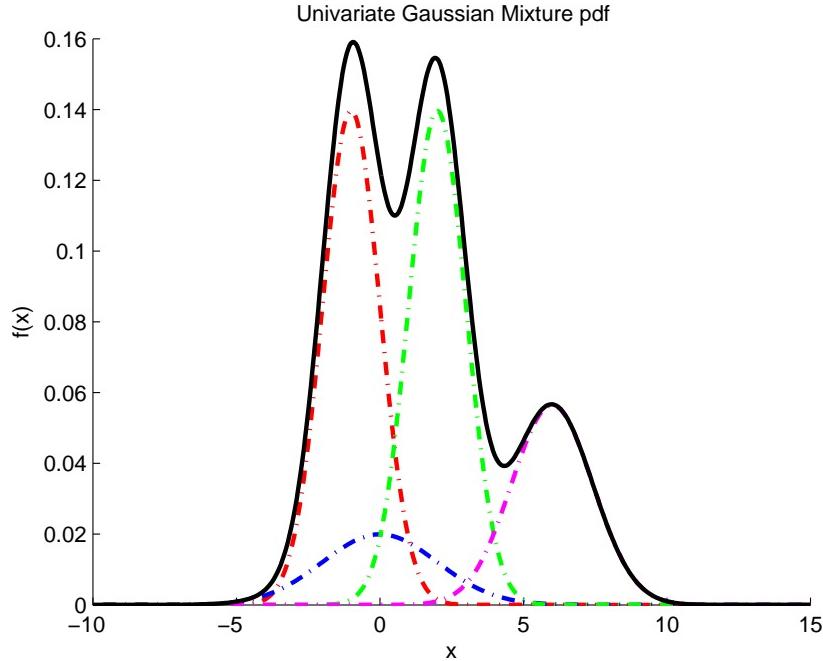


Figure 2.3: An illustration of a four-component univariate Gaussian mixture pdf (the solid line). Note that mixture components (represented by the dash-dotted traces) are scaled by their respective mixture weights in this graphic.

equations.

$$\begin{aligned}
 \boldsymbol{\mu} &= \sum_{i=1}^M p_i \boldsymbol{\mu}_i \\
 \mathbf{P} &= \sum_{i=1}^M p_i (\mathbf{P}_i + \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T) - \boldsymbol{\mu} \boldsymbol{\mu}^T \\
 &= \sum_{i=1}^M p_i [\mathbf{P}_i + (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T]
 \end{aligned} \tag{2.25}$$

Merging components of a target state Gaussian mixture pdf is one of two kinds of mixture reduction actions which will be used in this thesis. In Chapter 3 of [31], Salmond derived the equations for the merged mixture component weight, mean vector, and covariance matrix resulting from merging two or more components of the original multivariate Gaussian mixture pdf. These equations were derived under the constraint that the overall mean and covariance of the original mixture is pre-

served [31]. The new merged component parameters are

$$\begin{aligned}
p_m &= \sum_{i \in \mathcal{I}} p_i \\
\boldsymbol{\mu}_m &= \frac{1}{p_m} \sum_{i \in \mathcal{I}} p_i \boldsymbol{\mu}_i \\
\mathbf{P}_m &= \frac{1}{p_m} \sum_{i \in \mathcal{I}} p_i (\mathbf{P}_i + \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T) - \boldsymbol{\mu}_m \boldsymbol{\mu}_m^T
\end{aligned} \tag{2.26}$$

where $i \in \mathcal{I}$ indicates that the summation is taken only over those components that are merged and the subscript “m” is used to differentiate the *merged* component parameters from the others. As an example, if mixture components 1 and 2 are merged, then the resulting merged-component weight, mean vector, and covariance matrix would be (as derived by Williams in [38])

$$\begin{aligned}
p_{12} &= p_1 + p_2 \\
\boldsymbol{\mu}_{12} &= \frac{1}{(p_1 + p_2)} (p_1 \boldsymbol{\mu}_1 + p_2 \boldsymbol{\mu}_2) \\
\mathbf{P}_{12} &= \frac{1}{(p_1 + p_2)} \left[p_1 \mathbf{P}_1 + p_2 \mathbf{P}_2 + \frac{p_1 p_2}{p_1 + p_2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \right].
\end{aligned}$$

Deleting a mixture component is the second kind of mixture reduction action that will be used in this thesis. If a component of a target state Gaussian mixture pdf is deleted, then all one needs to do is ensure that the reduced set of mixture weights is re-normalized so that they sum to one.

2.4 Bayesian Approaches for Kinematics Model Parameter Uncertainty

In the previous sections, uncertainty was represented by the white Gaussian noise process vectors $\mathbf{w}_d(k - 1)$ and $\mathbf{v}(k)$ and the initial conditions on the state random process vector in Equation (2.5) in the discrete-time system kinematics and measurement models. This section introduces a new, realistic source of uncertainty encountered in target tracking. Two recursive Bayesian approaches are formulated depending on how one chooses to represent the uncertainty.

Consider the discrete-time linear system dynamics model

$$\mathbf{x}(k) = \Phi(k, k - 1)\mathbf{x}(k - 1) + \mathbf{G}_d\mathbf{w}_d(k - 1) \quad (2.4)$$

where $\mathbf{w}_d(k)$ is zero-mean with covariance \mathbf{Q}_d . This equation represents the equivalent discrete-time model of the continuous-time system describing the motion of a target over time. If this model does not adequately describe the target's motion, then the state vector estimated by a Kalman filter based on this model will likely be very inaccurate. The problem is not the Kalman filter, but the assumption that the system dynamics model sufficiently describes the target's motion over the time of interest. One remedy for this situation is to *use more than one model* when estimating the state random process vector. The underlying assumption is that at least one of the employed models is an adequate characterization of the target's kinematics. As a designer, this begs the question, "How can this be done?".

Assuming that the form of the linear system dynamics model in Equation (2.4) is correct, the answer to this question begins with the model matrices $\Phi(k, k - 1)$, \mathbf{G}_d , and \mathbf{Q}_d . Fundamentally, system dynamics models are defined by the elements of $\Phi(k, k - 1)$, \mathbf{G}_d , and \mathbf{Q}_d . One way to represent the uncertainty mathematically in the kinematics model is to assign these elements to random quantities. If these random matrix elements are placed into a vector, then the vector is either a continuous random vector (i.e., their values are unknown *constants*) denoted by \mathbf{M} (for "non-switching" models) or a continuous random *process* vector $\mathbf{M}(k)$ (for "switching"

models). The difference between non-switching and switching models is that the latter type includes a temporal dependence which admits the possibility that the target can assume different dynamics models over the tracking time of interest. However, a non-switching model-based Bayesian solution may also be used to track a target which exhibits various dynamics over a time interval of interest by including the appropriate modifications to enable model switching [23].

In general, the sample space of the random vectors representing the elements of the $\Phi(k, k - 1)$, \mathbf{G}_d , and \mathbf{Q}_d matrices is the n -dimensional real vector space \mathbb{R}^n : $\mathbf{M}, \mathbf{M}(k) \in \mathbb{M} \subset \mathbb{R}^n$ [22]. Conceptually, the continuous sample space \mathbb{M} contains *every* possible model that has the form of Equation (2.4). However, since the sample space includes every possible model, an uncountably infinite number of matrix element combinations is possible, and the Bayesian solution is not well-suited to real-time application [22].

Despite this obstacle, two recursive Bayesian approaches suitable for real-time implementation may be formulated by making a finite discretization of the continuous sample space, \mathbb{M} , along with other modifications. Subsection 2.4.1 introduces the recursive Bayesian solution for non-switching models and the *ad hoc* modifications necessary for this solution to accommodate target maneuvering. Subsection 2.4.2 presents the recursive Bayesian solution for switching models and three common approximations required to make this solution computationally tractable.

2.4.1 Non-Switching Models. For non-switching models, the kinematics of the target is assumed to be adequately modeled by at least one model in the continuous sample space, \mathbb{M} , for all times of interest (e.g., a plane travels according to model A for the observation times of interest). In most practical target tracking applications, this constraint is unrealistic since, for example, an aircraft could move at a constant velocity for a period of time and then perform some type of maneuver while in the surveillance region of an enemy tracking system. However, two *ad hoc* modifications to this approach allow the models to switch over time and accommodate maneuvering

targets [4, 23]. In this thesis, a non-switching model Bayesian solution that incorporates the *ad hoc* modifications to enable model switching is called a Multiple Model Adaptive Estimation (MMAE) algorithm.

The Bayesian solution for non-switching models is derived in this section. Initially, the model sample space \mathbb{M} is the continuous vector space \mathbb{R}^n , but this space is then discretized such that $\mathbb{M} = \{\mathbf{M}_i\}_1^{N_f}$, where N_f is the number of models (and thus the number of elemental filters in the non-switching model algorithm), to obtain a real-time solution [22]. Care must be taken when choosing the discretization of the original sample space so that at least one model in the set $\{\mathbf{M}_i\}_1^{N_f}$ adequately describes the kinematics of the target for all times of interest.

The derivation begins by modifying Equation (2.15) to include the *random* vector \mathbf{M} as a quantity to be estimated by inserting it to the left of the conditioning symbol. Assuming the joint pdf of $\mathbf{x}(k)$, \mathbf{Z}^k , and \mathbf{M} exists, the recursive Bayesian solution is

$$\begin{aligned} f(\mathbf{x}(k), \mathbf{M} | \mathbf{Z}^k) &= \frac{f(\mathbf{x}(k), \mathbf{M}, \mathbf{Z}^k)}{f(\mathbf{Z}^k)} \\ &= \frac{f(\mathbf{x}(k) | \mathbf{M}, \mathbf{Z}^k) f(\mathbf{M} | \mathbf{Z}^k) f(\mathbf{Z}^k)}{f(\mathbf{Z}^k)} \\ &= f(\mathbf{x}(k) | \mathbf{M}, \mathbf{Z}^k) f(\mathbf{M} | \mathbf{Z}^k) \end{aligned} \quad (2.27)$$

where the first pdf is the Kalman filter solution conditioned on a given model (a Gaussian pdf) and the second pdf is of the model conditioned on knowledge of the measurement history [22, 23]. The focus of the remainder of the derivation is on the second conditional pdf, $f(\mathbf{M} | \mathbf{Z}^k)$.

The model conditional pdf in Equation (2.27) evaluates to

$$\begin{aligned}
f(\mathbf{M}|\mathbf{Z}^k) &= \frac{f(\mathbf{M}, \mathbf{z}(k), \mathbf{Z}^{k-1})}{f(\mathbf{z}(k), \mathbf{Z}^{k-1})} \\
&= \frac{f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) f(\mathbf{M}|\mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})}{\int_{\mathbb{M}} f(\mathbf{z}(k), \mathbf{M}, \mathbf{Z}^{k-1}) d\mathbf{M}} \\
&= \frac{f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) f(\mathbf{M}|\mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})}{\int_{\mathbb{M}} f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) f(\mathbf{M}|\mathbf{Z}^{k-1}) d\mathbf{M} f(\mathbf{Z}^{k-1})} \\
&= \frac{f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) f(\mathbf{M}|\mathbf{Z}^{k-1})}{\int_{\mathbb{M}} f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) f(\mathbf{M}|\mathbf{Z}^{k-1}) d\mathbf{M}}
\end{aligned} \tag{2.28}$$

where the denominator is seen as the marginal pdf of the joint random vectors \mathbf{Z}^k and \mathbf{M} , integrating out the dependence on \mathbf{M} . In general, the integral in the denominator will require a computationally costly numerical solution and will likely prohibit the use of Equation (2.28) in an online implementation [22]. This problem is overcome by a finite discretization of \mathbb{M} [22].

Now, instead, let $\mathbb{M} = \{\mathbf{M}_i\}_1^{N_f}$. Then \mathbf{M} becomes a discrete random vector and its pdf may be written in terms of the probability mass function (pmf)

$p(\mathbf{M}_j) = \mathbf{P}[\mathbf{M} = \mathbf{M}_j]$ as¹¹

$$\begin{aligned} f(\mathbf{M}|\mathbf{Z}^k) &= \sum_{i=1}^{N_f} p(\mathbf{M} = \mathbf{M}_i | \mathbf{Z}^k) \delta(\mathbf{M} - \mathbf{M}_i) \\ &= \sum_{i=1}^{N_f} \mu_i(k) \delta(\mathbf{M} - \mathbf{M}_i) \end{aligned} \quad (2.29)$$

where $\mu_i(k) \triangleq p(\mathbf{M} = \mathbf{M}_i | \mathbf{Z}^k)$ is the hypothesis conditional probability or *mode probability* [4, 22, 23]. This quantity represents the probability that model i is correct, given the observed measurements [4, 22]. The mode probabilities are constrained by

$$\mu_i(k) \geq 0 \quad , i = 1, \dots, N_f, \quad \text{and} \quad \sum_{i=1}^{N_f} \mu_i(k) = 1$$

which will hold true (and will subsequently be shown) as long as $\mu_i(0) \geq 0, \forall i$ and $\sum_{i=1}^{N_f} \mu_i(0) = 1$ [22].

¹¹The generalized definition of the pdf allows for a pdf representation of a discrete random variable by noting that

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k)$$

where $F_X(x)$ is the cumulative distribution function of the random variable X and $p_X(x_k)$ is the pmf of X [20]. The derivative of $F_X(x)$ is, by definition, $f_X(x)$, so the pdf of a discrete random variable is [20]

$$\frac{d}{dx} F_X(x) = f_X(x) = \sum_k p_X(x_k) \delta(x - x_k).$$

Substituting the expression above for $f(\mathbf{M}|\mathbf{Z}^k)$ and $f(\mathbf{M}|\mathbf{Z}^{k-1})$ in Equation (2.28) results in

$$\begin{aligned}
\sum_{i=1}^{N_f} \mu_i(k) \delta(\mathbf{M} - \mathbf{M}_i) &= \frac{f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) \sum_{i=1}^{N_f} \mu_i(k-1) \delta(\mathbf{M} - \mathbf{M}_i)}{\int_{\mathbb{M}} f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) \sum_{i=1}^{N_f} \mu_i(k-1) \delta(\mathbf{M} - \mathbf{M}_i) d\mathbf{M}} \\
&= \frac{f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) \sum_{i=1}^{N_f} \mu_i(k-1) \delta(\mathbf{M} - \mathbf{M}_i)}{\sum_{i=1}^{N_f} \mu_i(k-1) \int_{\mathbb{M}} f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) \delta(\mathbf{M} - \mathbf{M}_i) d\mathbf{M}} \\
&= \frac{f(\mathbf{z}(k)|\mathbf{M}, \mathbf{Z}^{k-1}) \sum_{i=1}^{N_f} \mu_i(k-1) \delta(\mathbf{M} - \mathbf{M}_i)}{\sum_{i=1}^{N_f} \mu_i(k-1) f(\mathbf{z}(k)|\mathbf{M}_i, \mathbf{Z}^{k-1})} \quad (2.30)
\end{aligned}$$

by the sifting property of the delta function. To find a particular mode probability, simply set both sides of the above equation to $\mathbf{M} = \mathbf{M}_j$ where \mathbf{M}_j is the model of interest. Applying this approach for $j = 1, \dots, N_f$ yields an expression for each mode probability [4, 22, 23, 33]:

$$\mu_j(k) = \frac{\mu_j(k-1) f(\mathbf{z}(k)|\mathbf{M}_j, \mathbf{Z}^{k-1})}{\sum_{i=1}^{N_f} \mu_i(k-1) f(\mathbf{z}(k)|\mathbf{M}_i, \mathbf{Z}^{k-1})}, \quad j = 1, \dots, N_f. \quad (2.31)$$

Equation (2.31) indicates that the second term in the numerator divided by the denominator is always less than one as long as the initial constraints cited above are met (the $\mu_j(0)$'s sum to one and each $\mu_j(0)$ is greater than or equal to zero). Consequently, the sum of all mode probabilities is one for any sample k .

The expression $f(\mathbf{z}(k)|\mathbf{M}_j, \mathbf{Z}^{k-1})$ is the conditional pdf of $\mathbf{z}(k)$ conditioned on the assumed mode and the observed prior measurement history. At sample k the

current measurement is observed (in fact, it is \mathbf{z}_k , which is a realization of the random process $\mathbf{z}(k)$ at sample k), and the pdf $f(\mathbf{z}(k)|\mathbf{M}_j, \mathbf{Z}^{k-1})$ becomes the likelihood function of mode j given by [4, 22]

$$\begin{aligned}
f(\mathbf{z}(k)|\mathbf{M}_j, \mathbf{Z}^{k-1})|_{\mathbf{z}(k)=\mathbf{z}_k} &= L(\mathbf{z}_k; \mathbf{r}_{k,j}, \mathbf{S}_j(k)) \\
&= \frac{\exp\left[-\frac{1}{2}\mathbf{r}_{k,j}^T \mathbf{S}_j^{-1}(k) \mathbf{r}_{k,j}\right]}{(2\pi)^{\frac{m}{2}} \sqrt{\det \mathbf{S}_j(k)}} \\
\mathbf{r}_{k,j} &= \mathbf{z}_k - \mathbf{H}_j \hat{\mathbf{x}}_j(k|k-1) \\
\mathbf{S}_j(k) &= \mathbf{H}_j \mathbf{P}_j(k|k-1) \mathbf{H}_j^T + \mathbf{R}_j(k). \tag{2.32}
\end{aligned}$$

In this notation j indicates the model number, $j = 1, \dots, N_f$, so that each quantity above corresponds to one of the N_f filters. Also, m is the dimension of the random measurement vector $\mathbf{z}(k)$ or, equivalently, the dimension of $\mathbf{r}_{k,j}$.

To complete the recursive Bayesian estimator derivation for the case of non-switching models, substitute (2.29) and (2.15) into (2.27),

$$\begin{aligned}
f(\mathbf{x}(k), \mathbf{M}|\mathbf{Z}^k) &= f(\mathbf{x}(k)|\mathbf{M}, \mathbf{Z}^k) \sum_{i=1}^{N_f} \mu_i(k) \delta(\mathbf{M} - \mathbf{M}_i) \\
&= \sum_{i=1}^{N_f} \mu_i(k) f(\mathbf{x}(k)|\mathbf{M}_i, \mathbf{Z}^k) \\
&= \sum_{i=1}^{N_f} \mu_i(k) \frac{f(\mathbf{z}(k)|\mathbf{x}(k), \mathbf{M}_i, \mathbf{Z}^{k-1})}{f(\mathbf{z}(k)|\mathbf{M}_i, \mathbf{Z}^{k-1})} \cdot \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}_i(k|k-1), \mathbf{P}_i(k|k-1)\} \\
&= \sum_{i=1}^{N_f} \mu_i(k) \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}_i(k|k), \mathbf{P}_i(k|k)\}. \tag{2.33}
\end{aligned}$$

Equation (2.33) has nice interpretations in the context of the Kalman filter derivation in Subsection 2.2.1 and the multivariate Gaussian mixture introduction in Section 2.3. The last two lines of (2.33) indicate that the time propagation and measurement update stages of the non-switching model algorithm are comprised of N_f single-model Kalman filter time propagation and measurement update stages. Thus, N_f Kalman filters are needed, and each filter recursively operates under its own state random process vector estimates. The last line of Equation (2.33) is a Gaussian mixture since it is a weighted sum of Gaussian pdfs and the constraints on $\mu_i(k)$ are the same constraints imposed on p_i . Each mixture component, which is the weighted output of a Kalman filter with a distinct system dynamics model, may be interpreted as corresponding to the hypothesis that model j is correct. In this sense, the Bayesian solution for kinematics model uncertainty evaluates hypotheses about which model in the design best matches the target kinematics given the observed measurements.

Figure 2.4 depicts a block diagram of the non-switching model Bayesian solution inferred from Equation (2.33). Notice that each filter operates recursively under its own mean and covariance estimates. The overall mean and covariance of the state vector are

$$\begin{aligned}\hat{\mathbf{x}}(k|k) &= \sum_{i=1}^{N_f} \mu_i(k) \hat{\mathbf{x}}_i(k|k) \\ \mathbf{P}(k|k) &= \sum_{i=1}^{N_f} \mu_i(k) [\mathbf{P}_i(k|k) + \hat{\mathbf{x}}_i(k|k) \hat{\mathbf{x}}_i(k|k)^T] - \hat{\mathbf{x}}(k|k) \hat{\mathbf{x}}(k|k)^T \\ &= \sum_{i=1}^{N_f} \mu_i(k) [\mathbf{P}_i(k|k) + (\hat{\mathbf{x}}_i(k|k) - \hat{\mathbf{x}}(k|k))(\hat{\mathbf{x}}_i(k|k) - \hat{\mathbf{x}}(k|k))^T]\end{aligned}\quad (2.34)$$

where the state random process vector mean estimate is given by $\hat{\mathbf{x}}(k|k)$ and the state random process vector covariance estimate is $\mathbf{P}(k|k)$.

As noted, the non-switching model Bayesian solution for kinematics model parameter uncertainty does not account for the realistic possibility of target maneuvering. This shortfall is evident in the recursive mode probability calculation given

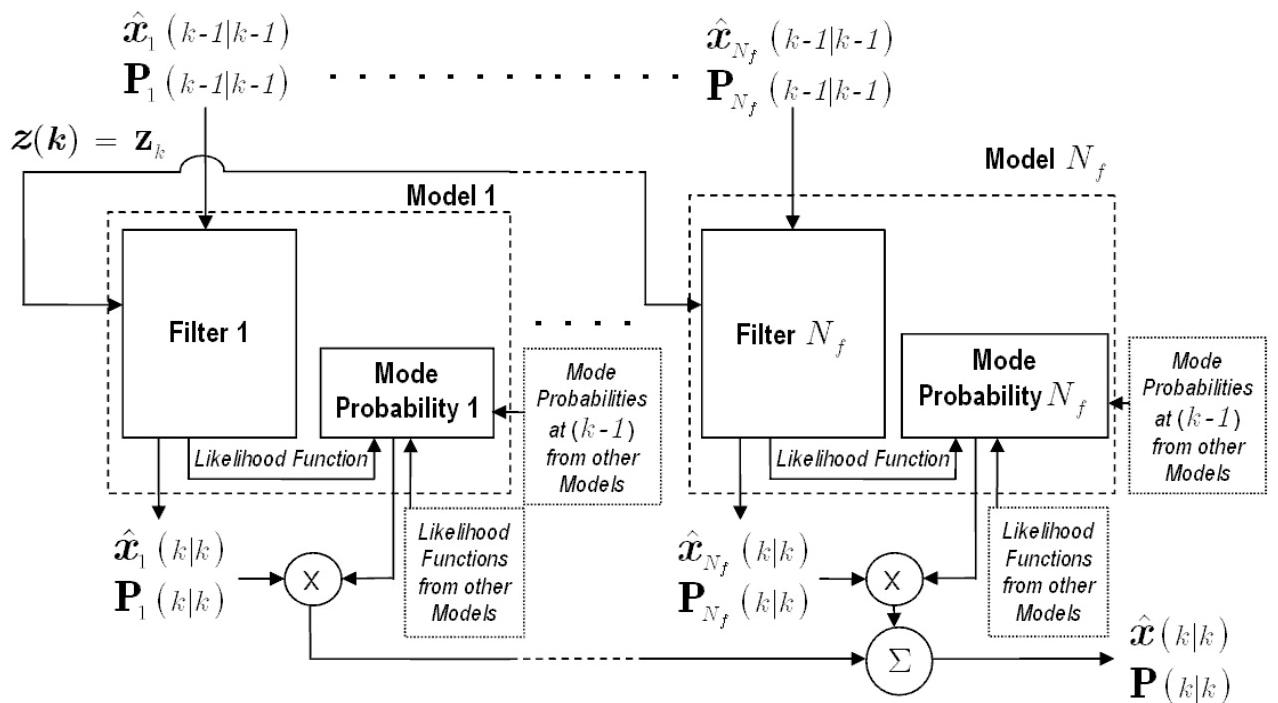


Figure 2.4: A block diagram of the non-switching multiple model algorithm. When certain *ad hoc* modifications are made to the non-switching model algorithm to enable model switching, then the algorithm is called MMAE.

in Equation (2.31). If any of the mode probabilities reach zero, then this mode probability will remain zero even if future measurements indicate (through the likelihood function calculation in Equation (2.32)) that the corresponding model is the best match to the current target dynamics. However, by introducing an artificial non-zero modal probability lower bound into Equation (2.31), this shortfall is overcome [4, 22, 23, 38]. Consider a scenario in which a target's dynamics over a long period of time were best represented by the first of two models so that the mode probability of the unfavorable filter is at the lower bound. Now imagine that the target assumes a new trajectory over an extended period of time which is best described by the second model. Then the observed measurements, through the likelihood function in Equation (2.32), would indicate that the second model is a more favorable match to the target's present dynamics. Since the modal probability of the second filter is at the lower bound and not zero, the modal probability will increase over time to favor the hypothesis that the target's motion is best described by the second model. A second *ad hoc* modification improves the response time of the algorithm to changes in target dynamics. Re-initializing the estimates of divergent filters allows the modal probability for unfavorable filters to increase more quickly in value if the observed measurements indicate that this filter is a good match. If the scalar quadratic form (Chi-square variable) $[\mathbf{r}_{k,j}^T \mathbf{S}_j^{-1}(k) \mathbf{r}_{k,j}]$ in Equation (2.32) is substantially greater than m , then that elemental filter can be declared divergent, and then be restarted with the state estimate in Equation (2.34) without the divergent elemental filter contributions included (and the mode probabilities rescaled so that the sum of these probabilities is one). MMAE implements *modal probability lower bounding* and *divergent filter re-initialization* to modify the non-switching model Bayesian solution for kinematics model uncertainty to accommodate maneuvering targets.

2.4.2 Switching Models. The switching model Bayesian solution for kinematics model uncertainty addresses the possibility of a maneuvering target up-front by mathematically modeling the uncertainty in the appropriate elements of the state

transition, noise input, and model noise process vector covariance matrices as random processes. It was shown in the previous subsection that allowing the model vector to have a continuous sample space would not likely lead to a real-time implementation, so the sample space of the model random process vector, $\mathbf{M}(k)$, is restricted to a finite discrete set $\{\mathbf{M}_i\}_1^{N_f}$ for all values of k [4]. Since $\mathbf{M}(k)$ is now discrete-valued, let $\mathbf{M}(k) = \mathbf{M}_{i_k}$. One drawback of representing the model vector as a random process is that the Bayesian solution needs to be approximated to obtain a computationally tractable algorithm. The recursive Bayesian solution in the presence of kinematics model parameter uncertainty is derived in this subsection, and three approximated solutions are developed in subsequent subsections.

As a starting point in finding a recursive Bayesian solution to this problem, consider the joint conditional pdf of the unknown target state random process vector and the unknown model random process vector $f(\mathbf{x}(k), \mathbf{M}(k)|\mathbf{Z}^k)$. After an application of the law of conditional probability for pdfs, this density becomes

$$f(\mathbf{x}(k), \mathbf{M}(k)|\mathbf{Z}^k) = f(\mathbf{x}(k)|\mathbf{M}(k), \mathbf{Z}^k) f(\mathbf{M}(k)|\mathbf{Z}^k) \quad (2.35)$$

which is similar to the joint conditional pdf in the non-switching model case (Equation (2.27)); that is the purpose of writing out Equation (2.35). However, it is shown in the remainder of this subsection that one must consider the model random process vector at all time instants through sample k (the *model history*) and not just $\mathbf{M}(k)$ itself to evaluate the pdfs on the right hand side of Equation (2.35) readily by means of a Kalman filter. Since $\mathbf{M}(k)$ is a discrete random vector at sample k , the second pdf is written as

$$f(\mathbf{M}(k)|\mathbf{Z}^k) = \sum_{i_k=1}^{N_f} p(\mathbf{M}(k) = \mathbf{M}_{i_k}|\mathbf{Z}^k) \delta(\mathbf{M}(k) - \mathbf{M}_{i_k}) \quad (2.36)$$

where the subscript on the summation index emphasizes that it corresponds to the discrete random vector $\mathbf{M}(k)$ at sample k . Since a recursive solution is sought so

that a Kalman filter may be applied to this problem, a relationship between $\mathbf{M}(k)$ and $\mathbf{M}(k-1), \mathbf{M}(k-2), \dots, \mathbf{M}(1)$ is needed. By using the fundamental definitions of discrete random processes and marginal pdfs [20], this pdf may be represented as (replacing $\mathbf{M}(k) = \mathbf{M}_{i_k}$ with simply \mathbf{M}_{i_k})

$$\begin{aligned} f(\mathbf{M}(k)|\mathbf{Z}^k) &= \sum_{i_k=1}^{N_f} \left(\sum_{i_{k-1}=1}^{N_f} \cdots \sum_{i_1=1}^{N_f} \right) p(\mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \dots, \mathbf{M}_{i_1} | \mathbf{Z}^k) \delta(\mathbf{M}(k) - \mathbf{M}_{i_k}) \\ &= \sum_{i_k=1}^{N_f} \left(\sum_{i_{k-1}=1}^{N_f} \cdots \sum_{i_1=1}^{N_f} \right) p(\mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_1^{k-1} | \mathbf{Z}^k) \delta(\mathbf{M}(k) - \mathbf{M}_{i_k}) \quad (2.37) \end{aligned}$$

which is, in fact, the marginal pdf of the joint pdf $f(\mathbf{M}(k), \dots, \mathbf{M}(1) | \mathbf{Z}^k)$. The summation notation i_{k-1} indicates that the summation index is for the discrete-valued random vector $\mathbf{M}(k-1) = \mathbf{M}_{i_{k-1}}$ which has a sample space of $\{\mathbf{M}_i\}_1^{N_f}$. A similar notational convention applies to the other subscripted i 's. Since each summation contains N_f terms and there are k summations, determining $f(\mathbf{M}(k)|\mathbf{Z}^k)$ requires evaluating $(N_f)^k$ terms. As time increases (i.e., k increases), the number of evaluations becomes unbounded and this solution becomes computationally intractable.

Still, the joint pmf $p(\mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_1^{k-1} | \mathbf{Z}^k)$ may be expanded using the law of conditional probability for pmfs in an attempt to reduce this expression to a recursive form:

$$\begin{aligned} p(\mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_1^{k-1} | \mathbf{Z}^k) &= \frac{f(\mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_1^{k-1}, \mathbf{z}(k), \mathbf{Z}^{k-1})}{f(\mathbf{z}(k), \mathbf{Z}^{k-1})} \quad (2.38) \\ &= \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \{\mathbf{M}_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}) p(\{\mathbf{M}_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})}. \end{aligned}$$

Observe that the measurement pdf $f(\mathbf{z}(k) | \cdot)$ as well as the model pmf $p(\mathbf{M}_{i_k} | \cdot)$ in the numerator of this equation are conditioned on the entire model history, $\{\mathbf{M}_{i_\ell}\}_1^{k-1}$.

Thus, a recursive solution would require knowledge of the model history over all previous samples.

However, these observations lead to two possible remedies to the problem of the exponential growth of evaluations. First, if \mathbf{M}_{i_k} is assumed to be a Markov process, then $p(\mathbf{M}_{i_k} | \{\mathbf{M}_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1})$ becomes $p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})$ (a transition probability from state $\mathbf{M}_{i_{k-1}}$ to state \mathbf{M}_{i_k}) and likewise for the other values of k ¹². By the definition of a Markov process, the transition probability $p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})$ solely depends on the previous state, so that the conditioning on \mathbf{Z}^{k-1} may be dropped; however, this conditioning will remain explicit in the notation for clarity. The second remedy is to limit the model histories for the measurement pdf above to the current sample or the current and previous time instants. Combining the Markov process assumption for \mathbf{M}_{i_k} and limiting the model histories for the measurement pdfs results in the Generalized pseudo-Bayesian (GPB) and Interacting Multiple Model (IMM) algorithms. Further approximations are necessary to produce practical real-time algorithms from GPB and IMM. These algorithms will be derived in the following subsections.

2.4.2.1 Generalized Pseudo-Bayesian-1 Algorithm. The GPB-1 algorithm uses the Markov process assumption for \mathbf{M}_{i_k} and limits the model history conditioning of the measurement pdf to the current sample k [4]. It also approximates the measurement history through sample $k - 1$ by the combined state random process vector mean and covariance estimates from the previous cycle. The algorithm requires N_f models, or filters, like the non-switching model algorithm, but model switches are enabled by the initial uncertainty modeling assumption given in Subsection 2.4.2. A derivation of the GPB-1 algorithm is contained in this subsection.

The GPB-1 derivation begins by using Equations (2.35) and (2.36) in their current form. Equation (2.38) is approximated using the Markov process assumption,

¹²By the definition of a Markov process, if $\mathbf{x}(k)$ is a discrete Markov process [20, 21], then

$$p(\mathbf{x}(k) | \mathbf{x}(k-1), \dots, \mathbf{x}(0)) = p(\mathbf{x}(k) | \mathbf{x}(k-1)).$$

and by discarding the conditioning on the previous model history in the measurement pdf:

$$p(\mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_{\ell=1}^{k-1} | \mathbf{Z}^k) \approx \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})}. \quad (2.39)$$

Substituting this expression into Equation (2.36) yields

$$\begin{aligned} f(\mathbf{M}(k) | \mathbf{Z}^k) &= \sum_{i_k=1}^{N_f} p(\mathbf{M}(k) = \mathbf{M}_{i_k} | \mathbf{Z}^k) \delta(\mathbf{M}(k) - \mathbf{M}_{i_k}) \\ &\approx \sum_{i_k=1}^{N_f} \delta(\mathbf{M}(k) - \mathbf{M}_{i_k}) \sum_{i_{k-1}=1}^{N_f} \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1}). \end{aligned} \quad (2.40)$$

The inner summation term is defined as the *mode probability* $\mu_{i_k}(k) \triangleq p(\mathbf{M}_{i_k} | \mathbf{Z}^k)$, and the term $p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})$ is $\mu_{i_{k-1}}(k-1)$, which leads to the recursion initially sought [4]. Additionally, $p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) \triangleq \tau_{i_k, i_{k-1}}$ is the *mode transition probability* which is simply the state transition probability for discrete Markov chains (this is a consequence of the Markov process approximation) [4]. The mode transition probability is chosen by the designer based on engineering insights.

Explicitly, the mode probabilities at sample k are given by

$$\mu_{i_k}(k) = \sum_{i_{k-1}=1}^{N_f} \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1) \quad (2.41)$$

for $i_k = 1, \dots, N_f$. The term in the denominator is a normalization factor and may be found by using the property of marginal pdfs/pmfss as

$$\begin{aligned}
f(\mathbf{z}(k) | \mathbf{Z}^{k-1}) &= \sum_{i_k=1}^{N_f} \frac{f(\mathbf{z}(k), \mathbf{M}_{i_k}, \mathbf{Z}^{k-1})}{f(\mathbf{Z}^{k-1})} \\
&= \sum_{i_k=1}^{N_f} \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})}{f(\mathbf{Z}^{k-1})} \\
&= \sum_{i_k=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \mathbf{Z}^{k-1}). \tag{2.42}
\end{aligned}$$

Now, the term $p(\mathbf{M}_{i_k} | \mathbf{Z}^{k-1})$ is

$$\begin{aligned}
p(\mathbf{M}_{i_k} | \mathbf{Z}^{k-1}) &= \sum_{i_{k-1}=1}^{N_f} \frac{f(\mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})}{f(\mathbf{Z}^{k-1})} \\
&= \sum_{i_{k-1}=1}^{N_f} \frac{p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1}) f(\mathbf{Z}^{k-1})}{f(\mathbf{Z}^{k-1})} \\
&= \sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1). \tag{2.43}
\end{aligned}$$

The final expression for the mode probabilities at sample k is now seen as

$$\mu_{i_k}(k) = \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}{\sum_{i_k=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)} \tag{2.44}$$

for $i_k = 1, \dots, N_f$, which ensures that the sum of all mode probabilities is one. The term $f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1})$ is simply the likelihood function given in Equation (2.32) but with i_k replacing j .

Substituting Equation (2.44) into Equations (2.40) and (2.35) yields

$$\begin{aligned}
f(\mathbf{x}(k), \mathbf{M}(k) | \mathbf{Z}^k) &= f(\mathbf{x}(k) | \mathbf{M}(k), \mathbf{Z}^k) f(\mathbf{M}(k) | \mathbf{Z}^k) \\
&\approx f(\mathbf{x}(k) | \mathbf{M}(k), \mathbf{Z}^k) \sum_{i_k=1}^{N_f} \mu_{i_k}(k) \delta(\mathbf{M}(k) - \mathbf{M}_{i_k}) \\
&= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{z}(k), \mathbf{Z}^{k-1}). \tag{2.45}
\end{aligned}$$

A final approximation is made by letting $\hat{\mathbf{x}}(k-1|k-1)$ and $\mathbf{P}(k-1|k-1)$ represent the information in the measurement history \mathbf{Z}^{k-1} so that the overall mean and covariance estimates from the previous cycle are propagated through each model; that is,

$$f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{z}(k), \mathbf{Z}^{k-1}) \approx f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{z}(k), \hat{\mathbf{x}}(k-1|k-1), \mathbf{P}(k-1|k-1)). \tag{2.46}$$

Finally, the target state pdf, under the GPB-1 assumption that \mathbf{M}_{i_k} is a Markov process and using the approximations (2.39) and (2.46), is

$$\begin{aligned}
f(\mathbf{x}(k), \mathbf{M}(k) | \mathbf{Z}^k) &\approx \sum_{i_k=1}^{N_f} \mu_{i_k}(k) f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{z}(k), \hat{\mathbf{x}}(k-1|k-1), \mathbf{P}(k-1|k-1)) \\
&= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}(k|k), \mathbf{P}(k|k)\} \tag{2.47}
\end{aligned}$$

with an overall mean and covariance given by

$$\begin{aligned}
\hat{\mathbf{x}}(k|k) &= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) \hat{\mathbf{x}}_{i_k}(k|k) \\
\mathbf{P}(k|k) &= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) [\mathbf{P}_{i_k}(k|k) + \hat{\mathbf{x}}_{i_k}(k|k) \hat{\mathbf{x}}_{i_k}(k|k)^T] - \hat{\mathbf{x}}(k|k) \hat{\mathbf{x}}(k|k)^T \\
&= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) [\mathbf{P}_{i_k}(k|k) + (\hat{\mathbf{x}}_{i_k}(k|k) - \hat{\mathbf{x}}(k|k)) (\hat{\mathbf{x}}_{i_k}(k|k) - \hat{\mathbf{x}}(k|k))^T]. \tag{2.48}
\end{aligned}$$

Equation (2.47) looks very similar to Equation (2.33) in that it represents a Gaussian mixture, but unlike that equation, the GPB-1 state vector pdf shows that the N_f models recursively operate on the *combined* state random process vector mean and covariance estimates. This fact is evident in the parameters of the Gaussian pdf in these two equations. For the non-switching model state vector pdf, the parameters of the Gaussian pdf inside the summation are $\hat{\mathbf{x}}_i(k|k)$ and $\mathbf{P}_i(k|k)$ where the subscript i indicates that these values are from the i^{th} filter. Thus the non-switching algorithm requires each filter to operate recursively on its own estimates. In contrast, the Gaussian pdf parameters in (2.47) are $\hat{\mathbf{x}}(k|k)$ and $\mathbf{P}(k|k)$ (the combined mean and covariance estimates of the target state vector), which show that each filter recursively operates according to the combined estimates of all filters. Additionally, Equation (2.47) includes the possibility of a model switch at any given time. That is, if the target changes from model \mathbf{M}_3 at time $k-1$ to model \mathbf{M}_5 at time k , then this model switch is characterized by the model transition probabilities, $\tau_{i_k, i_{k-1}}$, contained in $\mu_{i_k}(k)$. One potential drawback of GPB-1, and of any of the switching model algorithms (GPB-1, GPB-2, and IMM), is that the model transition probabilities, $\tau_{i_k, i_{k-1}}$, must be known. If these probabilities are not provided to a designer, then the designer must make an *ad hoc* assignment to their values. This last point demonstrates that *ad hoc* adjustments may be necessary for a practical implementation of the switching model algorithms, and that a practical Bayesian solution may require *ad hoc* modifications regardless of the initial assumption about the nature of the model vector (i.e., whether the model vector is represented as a random vector or a random process vector).

Figure 2.5 is a graphical representation of Equation (2.47). As previously noted, each filter runs under the combined target state vector estimates. Also note that the subscripts on the mean and covariance estimates of the state random process vector are the index values of the summation in Equation (2.47).

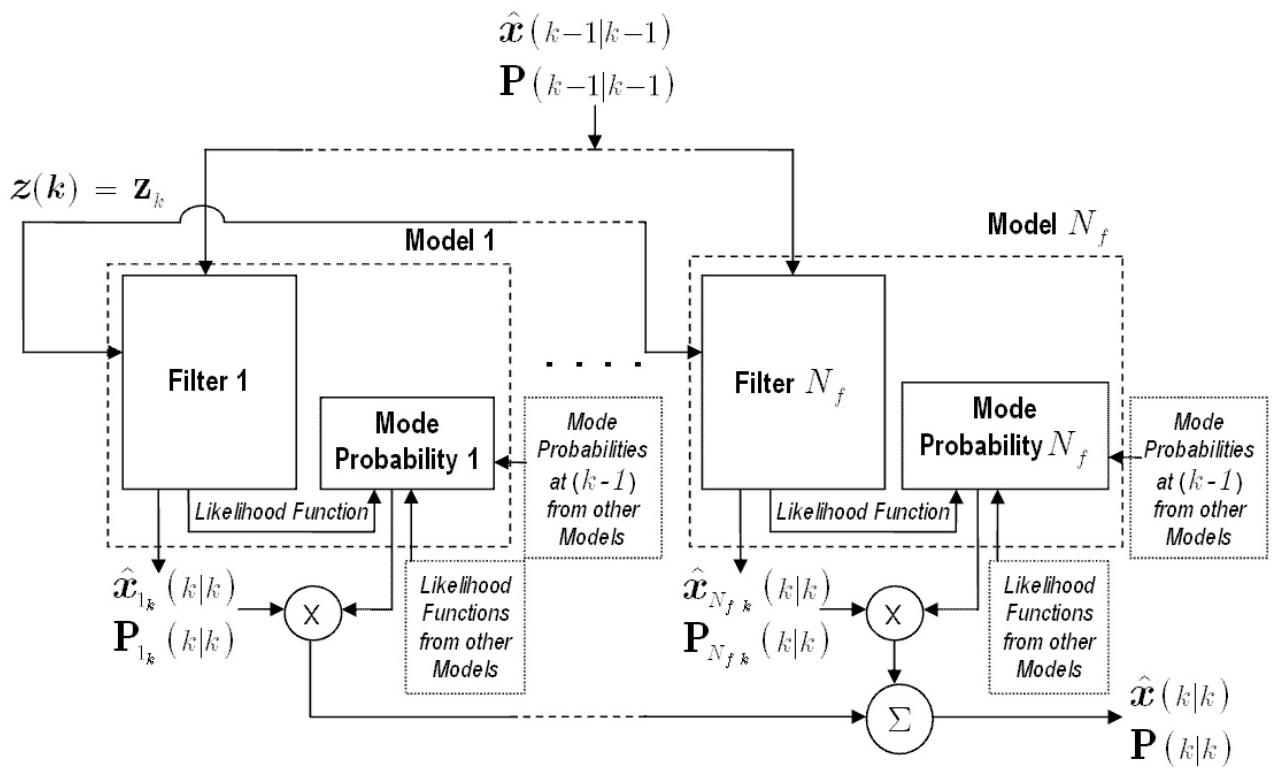


Figure 2.5: A block diagram of the GPB-1 algorithm.

2.4.2.2 Generalized Pseudo-Bayesian-2 Algorithm. Like GPB-1, the GPB-2 algorithm uses the Markov process assumption for \mathbf{M}_{i_k} , but now the measurement pdf is conditioned on the model vector at sample $k - 1$ in addition to the conditioning on the current model vector at sample k [4]. Since the measurement pdf conditioning includes the model vector at samples k and $k - 1$, the algorithm requires N_f^2 filters to operate. GPB-2 approximates the measurement and model histories through sample $k - 1$ by the weighted sum of the state random process vector mean and covariance estimates from the first set of filters (which will become clear when Equation (2.53) and Figure 2.6 are introduced). GPB-2 typically outperforms GPB-1, but at the expense of using N_f^2 filters as opposed to only N_f filters [4].

One assumption and two approximations to the switching model derivation are used to develop the GPB-2 algorithm:

1. \mathbf{M}_{i_k} is assumed to be a Markov process.
2. The conditional model pmf conditioned on the measurement history is approximated by conditioning the measurement pdf on the model vectors \mathbf{M}_{i_k} and $\mathbf{M}_{i_{k-1}}$ while discarding the model history at previous samples. Combining this approximation and condition 1 results in condition 2:

$$p(\mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_{\ell=1}^{k-1} | \mathbf{Z}^k) \approx \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})$$

In comparison, the corresponding GPB-1 approximation in Equation (2.39) only conditions the measurement pdf on the current model vector \mathbf{M}_{i_k} .

3. The measurement and target state pdfs are approximated by letting the state random process vector mean and covariance estimates from the previous sample represent the information contained in $\mathbf{M}_{i_{k-1}}$ and \mathbf{Z}^{k-1} . That is,

$$\begin{aligned}
f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) &\approx f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) \\
f(\mathbf{x}(k) | \mathbf{z}(k), \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) &\approx \\
&f(\mathbf{x}(k) | \mathbf{z}(k), \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1))
\end{aligned}$$

As before, a recursive Bayesian solution is desired, so (2.35) is used, but now $\mathbf{M}(k-1)$ is included in the target state pdf:

$$\begin{aligned}
f(\mathbf{x}(k), \mathbf{M}(k), \mathbf{M}(k-1) | \mathbf{Z}^k) = \\
f(\mathbf{x}(k) | \mathbf{M}(k), \mathbf{M}(k-1), \mathbf{Z}^k) f(\mathbf{M}(k-1) | \mathbf{M}(k), \mathbf{Z}^k) f(\mathbf{M}(k) | \mathbf{Z}^k). \quad (2.49)
\end{aligned}$$

Since $\mathbf{M}(k)$ and $\mathbf{M}(k-1)$ are discrete-valued random vectors, Equation (2.36) may be used in conjunction with the Markov assumption for the model random process vector (condition 1) to write the above equation as

$$\begin{aligned}
&f(\mathbf{x}(k), \mathbf{M}(k), \mathbf{M}(k-1) | \mathbf{Z}^k) \\
&= \sum_{i_k=1}^{N_f} \sum_{i_{k-1}=1}^{N_f} f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^k) p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) p(\mathbf{M}_{i_k} | \mathbf{Z}^k) \\
&= \sum_{i_k=1}^{N_f} \sum_{i_{k-1}=1}^{N_f} f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^k) p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \mu_{i_k}(k) \\
&= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) \sum_{i_{k-1}=1}^{N_f} f(\mathbf{x}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^k) p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \quad (2.50)
\end{aligned}$$

where $p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k)$ represents the *merging probabilities* and $\mu_{i_k}(k) = p(\mathbf{M}_{i_k} | \mathbf{Z}^k)$ is the mode probability [4]. Using conditions 1, 2, and 3, the merging probabilities

are given by (the overset number above the approximation symbols indicates which of the three conditions are used)

$$p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) = \quad (2.51)$$

$$\begin{aligned} & \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \{\mathbf{M}_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \{\mathbf{M}_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}) p(\{\mathbf{M}_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} \\ & \stackrel{1,2}{\approx} \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1}) \\ & = \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1) \\ & = \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}{\sum_{i_k=1}^{N_f} \sum_{i_{k-1}=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)} \\ & \stackrel{3}{\approx} \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}{\sum_{i_k=1}^{N_f} \sum_{i_{k-1}=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)} \end{aligned}$$

where $\tau_{i_k, i_{k-1}}$ is the mode transition probability given by $p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1})$ (as in Equation (2.40)) and $\mu_{i_{k-1}}(k-1)$, as given in Equation (2.41), is the previous sample mode probability, $p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})$. When the measurement is available, the measurement pdf of this equation, $f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1))$, becomes the likelihood function (2.32) for each of the i_k models with the appropriate changes

to the parameters. In a similar fashion, the current mode probabilities are given by

$$\mu_{i_k}(k) = p(\mathbf{M}_{i_k} | \mathbf{Z}^k) \quad (2.52)$$

$$\begin{aligned} & \sum_{i_{k-1}=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1) \\ & \approx \frac{\sum_{i_k=1}^{N_f} \sum_{i_{k-1}=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}{\sum_{i_k=1}^{N_f} \sum_{i_{k-1}=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}. \end{aligned}$$

Finally, after applying condition 3, the state random process vector pdf is

$$\begin{aligned} & f(\mathbf{x}(k), \mathbf{M}(k), \mathbf{M}(k-1) | \mathbf{Z}^k) \stackrel{3}{\approx} \sum_{i_k=1}^{N_f} p(\mathbf{M}_{i_k} | \mathbf{Z}^k) \\ & \cdot \sum_{i_{k-1}=1}^{N_f} f(\mathbf{x}(k) | \mathbf{z}(k), \mathbf{M}_{i_k}, \hat{\mathbf{x}}_{i_{k-1}}(k-1|k-1), \mathbf{P}_{i_{k-1}}(k-1|k-1)) p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \\ & = \sum_{i_k=1}^{N_f} \mu_{i_k}(k) \sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}_{i_{k-1}}(k|k), \mathbf{P}_{i_{k-1}}(k|k)\}. \end{aligned} \quad (2.53)$$

The inner summation is a Gaussian mixture with mean and covariance given by Equation (2.25). Once the inner summation is evaluated, then the outer summation produces another Gaussian mixture. That is, first compute

$$\begin{aligned} \hat{\mathbf{x}}_{i_k}(k|k) &= \sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \hat{\mathbf{x}}_{i_{k-1}}(k|k) \\ \mathbf{P}_{i_k}(k|k) &= \sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \left[\mathbf{P}_{i_{k-1}}(k|k) + \right. \\ &\quad \left. (\hat{\mathbf{x}}_{i_{k-1}}(k|k) - \hat{\mathbf{x}}_{i_k}(k|k))(\hat{\mathbf{x}}_{i_{k-1}}(k|k) - \hat{\mathbf{x}}_{i_k}(k|k))^T \right], \end{aligned} \quad (2.54)$$

then calculate the GPB-2 overall mean and covariance estimates as

$$\begin{aligned}\hat{\mathbf{x}}(k|k) &= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) \hat{\mathbf{x}}_{i_k}(k|k) \\ \mathbf{P}(k|k) &= \sum_{i_k=1}^{N_f} \mu_{i_k}(k) [\mathbf{P}_{i_k}(k|k) + (\hat{\mathbf{x}}_{i_k}(k|k) - \hat{\mathbf{x}}(k|k))(\hat{\mathbf{x}}_{i_k}(k|k) - \hat{\mathbf{x}}(k|k))^T].\end{aligned}\quad (2.55)$$

Figure 2.6 depicts a block diagram of the GPB-2 algorithm. At the beginning of a processing cycle, each of the merged estimates are input into N_f filters and the same measurements are fed to each filter. The superscripts on the mean and covariance estimates at the outputs of the N_f^2 filters correspond to the indices of the inner summation in Equation (2.53). Once the inner summation is calculated for each i_k of the outer summation, the estimates are *merged* after being scaled by the merging probabilities $p(\mathbf{M}_{i_{k-1}}|\mathbf{M}_{i_k}, \mathbf{Z}^k)$. Finally, the merged estimates are scaled by the mode probabilities $\mu_{i_k}(k)$ and combined via Equation (2.55) into the overall mean and covariance estimate of the state random process vector.

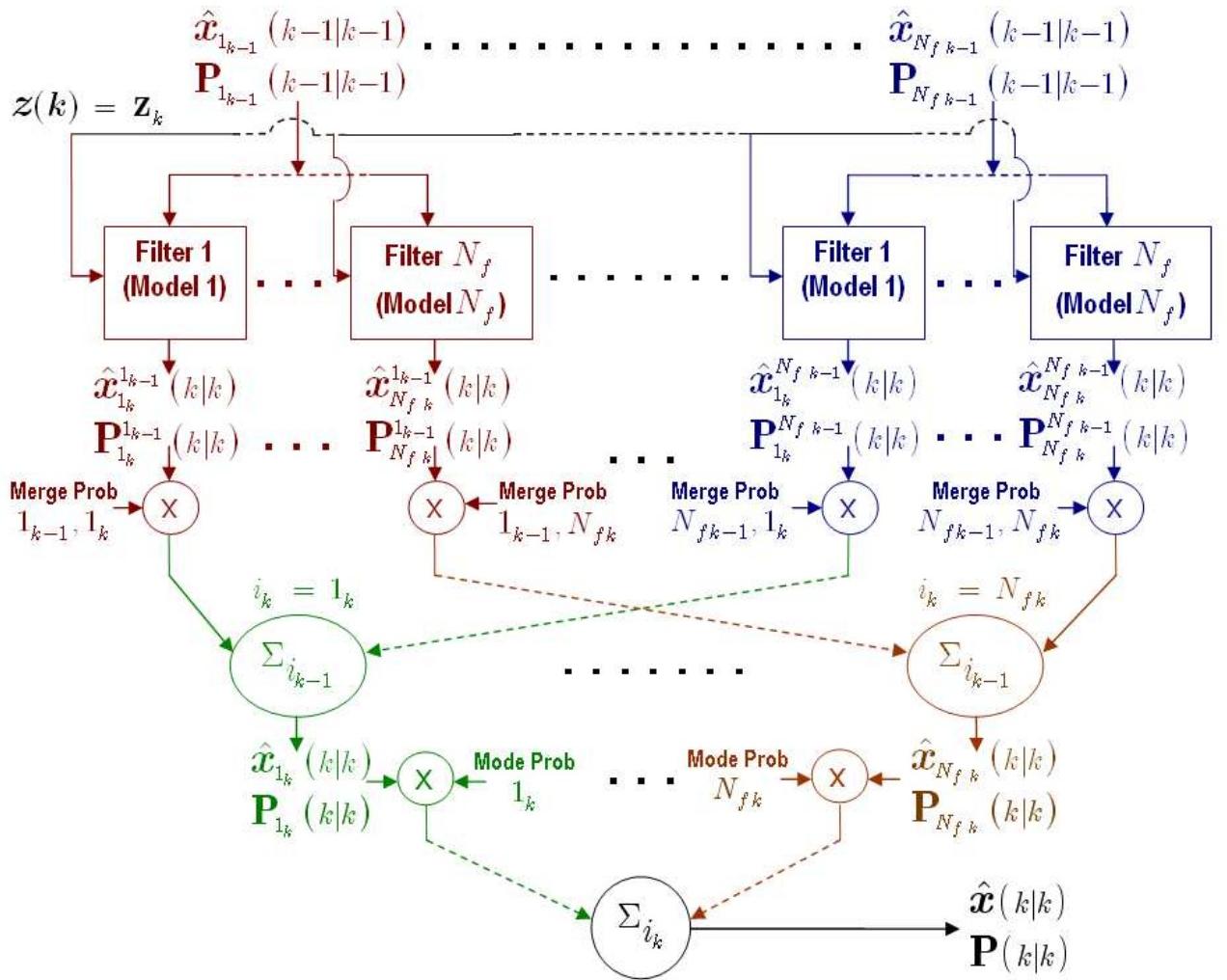


Figure 2.6: A notional block diagram of the GPB-2 algorithm.

2.4.2.3 Interacting Multiple Model Algorithm. The IMM algorithm achieves the performance of GPB-2 but uses only the same number of filters as GPB-1 [4, 38]. For these reasons IMM is the preferred approximation to the switching multiple model approach.

In [4], IMM is derived starting from the GPB-1 assumptions/approximations and later incorporating condition 3 from the GPB-2 subsection. This approach results in an algorithm in which the merging probabilities scale the estimate of the state random process vector from the previous cycle before beginning the next cycle of the algorithm. In [38] and [33], the derivation of IMM incorporates scaling by the merging probabilities after the end of each cycle¹³. This form of the derivation emphasizes that the IMM algorithm reduces to the non-switching algorithm if estimate merging does not take place (skipping ahead to Figure 2.7, if one removes the ‘‘Model Estimate Merging’’ block and feeds the previous estimates into the appropriate filters, then the IMM block diagram is essentially the same as the non-switching model block diagram in Figure 2.4). Mathematically, it can be shown that IMM reduces to the non-switching model solution when $p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k)$ is replaced by a Kronecker delta function, $\delta_{i_k i_{k-1}}$, for each pair of index values [33].

This subsection includes a third form of the IMM algorithm derivation which emphasizes two points not highlighted in [4, 33, 38]. First, IMM may be viewed as a Gaussian mixture reduction approximation of GPB-2 to decrease the number of filters from N_f^2 to N_f . The second point is more subtle than the first. The claim that IMM provides performance similar to GPB-2 is evident when the IMM algorithm is derived directly from GPB-2. Therefore, the third derivation of the IMM algorithm begins with GPB-2.

The inner summation of the GPB-2 target state pdf in Equation (2.53) is a Gaussian mixture. If the mixture is approximated by a *single* Gaussian pdf with the same overall mean and covariance as the original Gaussian mixture, then the num-

¹³Whether merging occurs at the beginning or the end of a process cycle is irrelevant. Both methods are theoretically equivalent.

ber of filters is reduced from N_f^2 to N_f (only the outer summation remains). In this way, IMM may be derived from GPB-2 by including this Gaussian mixture reduction approximation to the assumption/approximation conditions listed in Subsection 2.4.2.2.

Starting with the GPB-2 state random process vector pdf given in Equation (2.53), the IMM target state pdf is

$$\begin{aligned}
& f(\mathbf{x}(k), \mathbf{M}(k), \mathbf{M}(k-1) | \mathbf{Z}^k) \\
&= f(\mathbf{x}(k) | \mathbf{M}(k), \mathbf{M}(k-1), \mathbf{Z}^k) f(\mathbf{M}(k-1) | \mathbf{M}(k), \mathbf{Z}^k) f(\mathbf{M}(k) | \mathbf{Z}^k) \\
&\stackrel{GPB-2}{\approx} \sum_{i_k=1}^{N_f} p(\mathbf{M}_{i_k} | \mathbf{Z}^k) \sum_{i_{k-1}=1}^{N_f} \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}_{i_{k-1}}(k|k), \mathbf{P}_{i_{k-1}}(k|k)\} p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k) \\
&\stackrel{IMM}{\approx} \sum_{i_k=1}^{N_f} p(\mathbf{M}_{i_k} | \mathbf{Z}^k) \mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}_{i_k}(k|k), \mathbf{P}_{i_k}(k|k)\}. \tag{2.56}
\end{aligned}$$

The overset text on the first approximation, “GPB-2,” indicates that the assumption/approximation conditions in Subsection 2.4.2.2 are invoked in the approximation. Likewise, the overset text on the second approximation indicates that the IMM single Gaussian pdf mixture reduction approximation is used. The IMM target state pdf appears to be the same as that for GPB-1 in Equation (2.47), but it fundamentally differs from the target state pdf of GPB-1 since the merging probability terms, $p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k)$, are embedded in $\mathcal{N}\{\mathbf{x}(k); \hat{\mathbf{x}}_{i_k}(k|k), \mathbf{P}_{i_k}(k|k)\}$. Thus the performance of IMM is expected to be closer to that of GPB-2 than that of GPB-1. The parameters of the pdf, $\hat{\mathbf{x}}_{i_k}(k|k)$ and $\mathbf{P}_{i_k}(k|k)$, are given by Equation (2.54). However, the equation for $p(\mathbf{M}_{i_{k-1}} | \mathbf{M}_{i_k}, \mathbf{Z}^k)$ is calculated in a different manner than seen in Equation (2.51).

The Bayes expansion of the merging probability $p(\mathbf{M}_{i_{k-1}}|\mathbf{M}_{i_k}, \mathbf{Z}^k)$ differs from that of GPB-2 by conditioning on the current measurement history (through sample k) instead of the previous measurement history (through sample $k-1$). This modification enables merging of the previous cycle estimates *prior* to beginning the next cycle. To see this point, consider the expansion of $p(\mathbf{M}_{i_{k-1}}|\mathbf{M}_{i_k}, \mathbf{Z}^k)$ using Bayes' rule and marginal probability:

$$\begin{aligned}
p(\mathbf{M}_{i_{k-1}}|\mathbf{M}_{i_k}, \mathbf{Z}^k) &= \frac{p(\mathbf{M}_{i_k}|\mathbf{M}_{i_{k-1}}, \mathbf{Z}^k) p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^k)}{p(\mathbf{M}_{i_k}|\mathbf{Z}^k)} \\
&= \frac{p(\mathbf{M}_{i_k}|\mathbf{M}_{i_{k-1}}, \mathbf{Z}^k) p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^k)}{\sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_k}|\mathbf{M}_{i_{k-1}}, \mathbf{Z}^k) p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^k)} \\
&= \frac{\tau_{i_k, i_{k-1}} p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^{k-1})}{\sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^{k-1})} \\
&= \frac{\tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}{\sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}. \tag{2.57}
\end{aligned}$$

Two insights were used to obtain the last line of this equation. First, $p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^k)$ is equivalent to $p(\mathbf{M}_{i_{k-1}}|\mathbf{Z}^{k-1})$ since the measurement at sample k has no impact on the model vector at sample $k-1$. Second, $p(\mathbf{M}_{i_k}|\mathbf{M}_{i_{k-1}}, \mathbf{Z}^k)$ is represented by the model transition probability, $\tau_{i_k, i_{k-1}}$. It seems counterintuitive that the probability of the current model would not depend on the current measurement. However, the Markov process assumption for \mathbf{M}_{i_k} imposes the condition that the transition probability is *only* dependent on the previous model state. Therefore, the conditioning on \mathbf{Z}^k in $p(\mathbf{M}_{i_k}|\mathbf{M}_{i_{k-1}}, \mathbf{Z}^k)$ is irrelevant under the Markov process assumption.

All that remains of the IMM algorithm derivation is to determine the mode probability, $\mu_{i_k}(k) \triangleq p(\mathbf{M}_{i_k}|\mathbf{Z}^k)$, and the likelihood function, $f(\mathbf{z}(k)|\mathbf{M}_{i_k}, \mathbf{Z}^{k-1})$. The

mode probability is

$$\begin{aligned}
\mu_{i_k}(k) &= p(\mathbf{M}_{i_k} | \mathbf{z}(k), \mathbf{Z}^{k-1}) \\
&= \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \mathbf{Z}^{k-1})}{f(\mathbf{z}(k) | \mathbf{Z}^{k-1})} \\
&= \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})}{\sum_{i_k=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_k} | \mathbf{Z}^{k-1})} \\
&= \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})}{\sum_{i_k=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} p(\mathbf{M}_{i_k} | \mathbf{M}_{i_{k-1}}, \mathbf{Z}^{k-1}) p(\mathbf{M}_{i_{k-1}} | \mathbf{Z}^{k-1})} \\
&= \frac{f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)}{\sum_{i_k=1}^{N_f} f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1}) \sum_{i_{k-1}=1}^{N_f} \tau_{i_k, i_{k-1}} \mu_{i_{k-1}}(k-1)} \tag{2.58}
\end{aligned}$$

(note that the mode probability for IMM is the same as that for GPB-1 in Equation (2.44)). Using Equation (2.32), the likelihood function of $f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1})$ given the realized observation $\mathbf{z}(k) = \mathbf{z}_k$ is

$$\begin{aligned}
f(\mathbf{z}(k) | \mathbf{M}_{i_k}, \mathbf{Z}^{k-1})|_{\mathbf{z}(k)=\mathbf{z}_k} &= L(\mathbf{z}_k; \mathbf{r}_{i_k, k}, \mathbf{S}_{i_k}(k)) \\
&= \frac{\exp\left[-\frac{1}{2}\mathbf{r}_{i_k, k}^T \mathbf{S}_{i_k}^{-1}(k) \mathbf{r}_{i_k, k}\right]}{(2\pi)^{\frac{m}{2}} \sqrt{\det \mathbf{S}_{i_k}(k)}} \\
\mathbf{r}_{i_k, k} &= \mathbf{z}_k - \mathbf{H}_{i_k} \hat{\mathbf{x}}_{i_k}(k | k-1) \\
\mathbf{S}_{i_k}(k) &= \mathbf{H}_{i_k} \mathbf{P}_{i_k}(k | k-1) \mathbf{H}_{i_k}^T + \mathbf{R}_{i_k}(k). \tag{2.59}
\end{aligned}$$

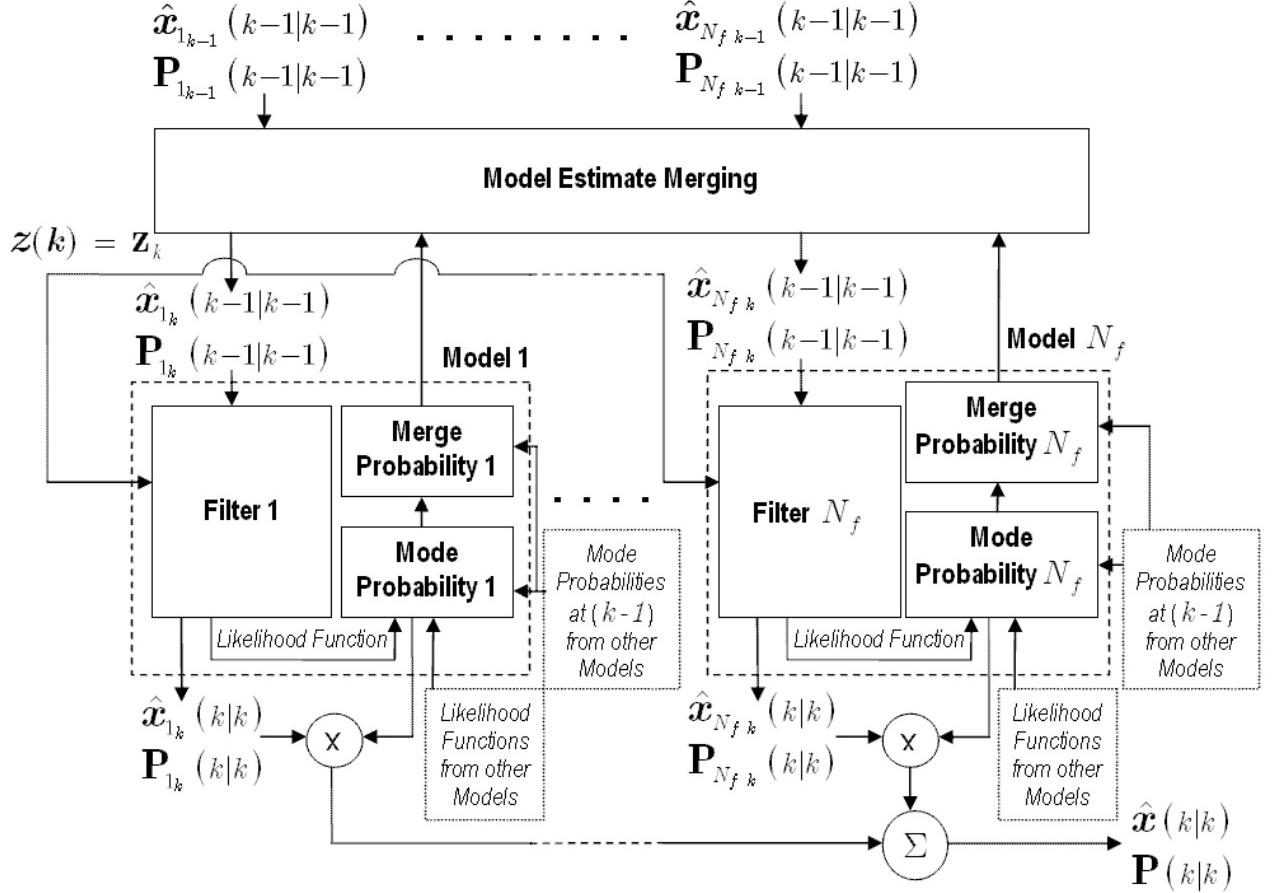


Figure 2.7: A block diagram of the IMM algorithm.

Figure 2.7 represents the IMM algorithm. At the beginning of each processing cycle, the N_f estimates of the last processing cycle are merged according to the merging probabilities of each model (filter). Merging occurs in the “Model Estimate Merging” block which functions as a mixture reduction algorithm, reducing the N_f -order Gaussian mixture to a single Gaussian pdf. The filters then operate on the merged estimates (note that the subscripts on the mean and covariance estimates of the state random process vector change from i_{k-1} to i_k , which corresponds to Equation (2.56)). After being scaled by the mode probability for each model, the outputs of the filters are combined into the overall estimates by the summation block.

2.4.3 Multiple Model Algorithms Summary. This section introduced the realistic problem of kinematics model parameter uncertainty encountered in target

tracking. This uncertainty may be represented by the unknown elements of the $\Phi(k, k - 1)$, \mathbf{G}_d , and \mathbf{Q}_d matrices, which are mathematically represented as random quantities and inserted into a parameter vector to be estimated. Since the model parameters are random quantities, Bayes estimation can be used to find a recursive Bayesian solution. It was quickly determined that the model parameters had to be limited to a discrete set of values to provide any hope of a real-time implementation. After limiting the continuous set of possible values to a finite discrete set, two fundamental assumptions about the random nature of the model parameters were made, leading to two Bayesian solutions. The first solution assumed that the model parameters were time-invariant so that they were represented by a *random vector* \mathbf{M} . This assumption led to the *non-switching* model recursive Bayesian solution. The fundamental drawback to representing the model parameters as random constants is that the resulting Bayesian solution presumes that the target travels according to only one model *for all time*. The second solution assumed that the model parameters were time-varying, and the model parameters were represented by a *random process vector* $\mathbf{M}(k)$. This assumption led to the *switching* model recursive Bayesian solution. The advantage of this method is that the target is *not* assumed to travel according to only one model for all time. Both the rigorous non-switching and switching model recursive Bayesian solutions are unsuitable for practical implementation in a target tracking system.

Ad hoc modifications and approximations to the rigorous non-switching and switching model solutions, respectively, led to practical implementations for target tracking systems. MMAE, which is based on the non-switching model recursive Bayesian solution, utilizes *ad hoc modal probability lower bounding* to enable model switching. Additionally, *filter re-initialization* is used to improve the response of the algorithm to changes in target dynamics. With these modifications to the non-switching model solution in place, MMAE is a suitable multiple model algorithm for use in real-time tracking of a maneuvering target. GPB-1, GPB-2, and IMM incorporate various approximations to the rigorous switching model Bayesian solution to

produce practical algorithms which inherently allow for target maneuvering. Of the three approximations, IMM provides performance approaching that of GPB-2 (which generally provides the best performance of the group), but with a computational burden on par with that of GPB-1 (and MMAE). In fact, IMM reduces to MMAE when the model transition probabilities are represented by Kronecker delta functions [33].

2.5 A Bayesian Approach for Measurement Origin Uncertainty

This section introduces another practical source of uncertainty: the origin of measurements. One aspect of the *data association* problem is depicted in Figure 2.8 in which fourteen measurements are observed when only two targets are known to exist. Assuming that each target cannot produce more than one return, and that feature information about each of the measurements (e.g., amplitude, phase, etc.) is not available, which two of the fourteen observed measurements originated from the two targets? This scenario is the typical data association problem in which the origin of all measurements generated by sensors in a scan must be determined. Another aspect of the data association problem allows for the possibility that, for example, all fourteen observed measurements were generated by fourteen new targets and the two existing targets were not detected.

Measurements may be broadly categorized into *true measurements* or *false-origin measurements*. True measurements include those belonging to hypothesized targets that were hypothesized in previous scans and are used for track continuation, or those belonging to potential new targets so that new tracks are initiated. A track is a state vector trajectory estimated from a set of measurements that have been associated with the same target over some number of scans [5]. False-origin measurements may arise from clutter, countermeasures, or false alarms. Clutter may be considered objects other than targets that create spurious returns. Countermeasures include decoys and jamming. False alarms are erroneous measurements caused by random sensor or environmental noise. The possibility of *missed measurements* also exists, which may occur when tracking low observable targets, for instance.

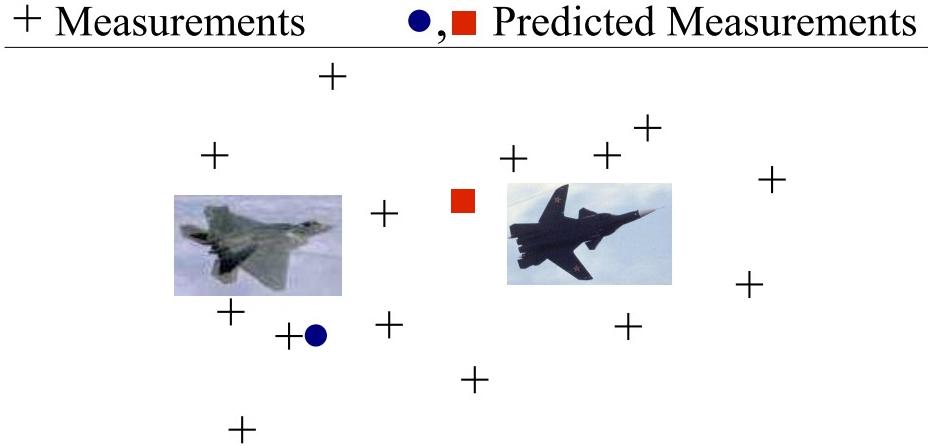


Figure 2.8: An illustration of one aspect of the data association problem. Given more observed measurements than targets, how does one update the target state random process vector estimates (adapted from [38])? The pictures of the F/A-22 and S-37 represent the true location of the targets.

Data association requires generating hypotheses about the origin of the measurements. These hypotheses are represented by *association events* which are formed by labeling the measurements according to one of two underlying assumptions used in practice. The first assumption is that the true number of targets is known exactly. In this case, one would associate measurements to known tracks or to false sources (clutter, countermeasures, or false alarms). Under this assumption, the tracker operates according to a *target-oriented* data association method [5]. The second underlying assumption, which seems more widely applicable, is that the true number of targets is not known. Under this assumption, measurements may be associated with existing tracks that were hypothesized in previous scans, potential new tracks, or false sources. This association method is called the *measurement-oriented* approach [27].

A commonly used method to reduce the number of association events is to place a *measurement gate* around the predicted measurement, $\hat{z}_j(k|k-1)$, for each existing target j that was hypothesized in a previous scan. Measurements that are within a measurement gate of an existing target at the current scan are hypothesized as potentially originating from that target, while measurements outside of the gate are

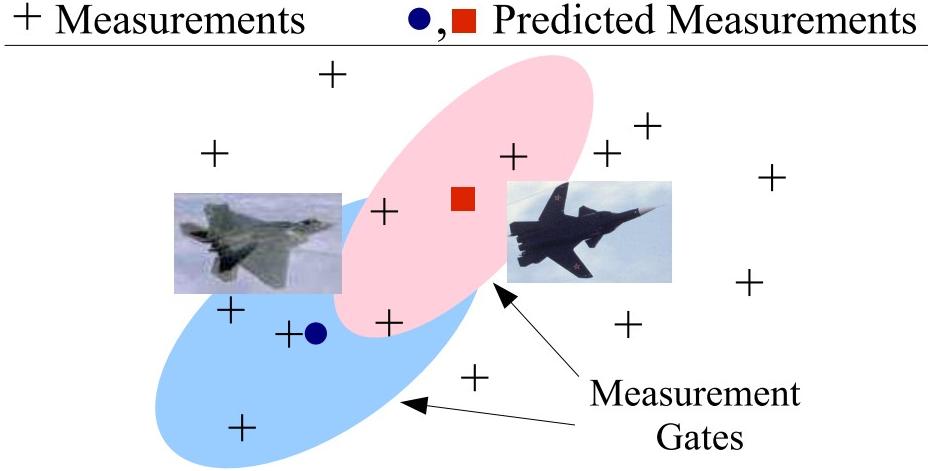


Figure 2.9: By using measurement gates based on the predicted measurement, $\hat{z}_j(k|k-1)$, for each existing target j that was hypothesized in a previous scan, the number of potential association events is decreased (adapted from [38]).

not considered as candidates for association to that target. Instead, measurements outside all of the measurement gates are hypothesized as false-origin measurements when the target-oriented data association method is used, or as false-origin or potential new target measurements when using the measurement-oriented data association approach. Figure 2.9 shows the predicted measurements for the existing targets that were hypothesized in a previous scan as a dot and a square, and the true target locations are indicated by the F/A-22 and S-37 images. Although there are several types of measurement gates [7], Figure 2.9 depicts two elliptical measurement gates centered about the predicted measurement for each target. The size of each gate is related to $\mathbf{S}_j(k)$, which is the covariance of the residual for each target j (see Subsection 2.2.1), and it may be specified in terms of the probability that the true target measurement falls within the gate, P_g [5]. Measurements outside the union of the measurement gates are considered too unlikely to have originated from the targets and, as a result, are hypothesized to have originated from other sources.

The union of all association events formed from labeling all of the measurements at sample k creates the discrete sample space of the association event discrete *random process vector* $\Theta(k) = \Theta_{i_k}$. Thus, as in the switching model case of the previous

section, the uncertainty about the true origin of the measurements is modeled by a discrete random process vector, and the mathematics of that section also apply to this section. The number of measurements received at each time instant is random in nature and, consequently, so is the size of the discrete sample space of Θ_{i_k} . As a result, the number of association events is also random.

As an illustration of generating association events at some sample k , consider the following example. A confirmed track is to be updated by two measurements at sample k . Furthermore, to simplify the example, measurement gating is not used, so either measurement may be associated with the confirmed track. Using the measurement-oriented data association method, the origin of each measurement may be the existing track that was hypothesized in a previous scan, a potential new track, or a false source. Assume that the measurements originate from at most one source and any, all, or none of the measurements may originate from potential new targets or false sources. Let H_{FT} , H_{DT} , and H_{NT} represent the hypotheses that one of the measurements originated from a false source, the existing track hypothesized in a previous scan (“ D ” is for “detected”), or a potential new track, respectively. Also, let $\mathbf{z}_{k,1}$ and $\mathbf{z}_{k,2}$ denote the two observed measurements. Under these conditions, one may generate all of the feasible association events by creating tables in which each column corresponds to a distinct hypothesis (H_{FT} , H_{DT} , or H_{NT}), each row corresponds to a measurement ($\mathbf{z}_{k,1}$ or $\mathbf{z}_{k,2}$), and each element contains a “1” or a “0” depending on whether a hypothesis/measurement pair is considered “true” or “false,” respectively.

Figure 2.10 depicts tables containing all feasible association events for the example given above. Since there is only one confirmed track and two measurements at sample k , the number of potential targets is three using the measurement-oriented data association method (the two measurements may be from two new targets). A “1” in the cell corresponding to hypothesis H_* and measurement $\mathbf{z}_{k,*}$ indicates that the hypothesis/measurement pair is true, while a “0” means that the hypothesis/measurement pair is not true. For instance, the third table in the first row shows that measurement $\mathbf{z}_{k,1}$ is associated with hypothesis H_{FT} , a false source, and measurement

	H_{FT}	H_{DT}	H_{NT}		H_{FT}	H_{DT}	H_{NT}		H_{FT}	H_{DT}	H_{NT}
$z_{k,1}$	0	1	0		0	1	0		1	0	0
$z_{k,2}$	1	0	0		0	0	1		0	1	0
	Θ_{1_k}				Θ_{2_k}				Θ_{3_k}		
	H_{FT}	H_{DT}	H_{NT}		H_{FT}	H_{DT}	H_{NT}		H_{FT}	H_{DT}	H_{NT}
$z_{k,1}$	0	0	1		1	0	0		0	0	1
$z_{k,2}$	0	1	0		0	0	1		1	0	0
	Θ_{4_k}				Θ_{5_k}				Θ_{6_k}		
	H_{FT}	H_{DT}	H_{NT}		H_{FT}	H_{DT}	H_{NT}		H_{FT}	H_{DT}	H_{NT}
$z_{k,1}$	1	0	0		0	0	1		0	0	1
$z_{k,2}$	1	0	0		0	0	1		0	0	0
	Θ_{7_k}				Θ_{8_k}						

Figure 2.10: Generated feasible association events at sample k for two measurements and one confirmed track. This representation of the discrete sample space of Θ_{i_k} contains eight association events or hypotheses.

$\mathbf{z}_{k,2}$ is associated with the confirmed target hypothesized in a previous scan H_{DT} . The true association event is assumed to be included in the sample space of the association event random process vector, Θ_{i_k} . Note that each row of every table must sum to one since a measurement cannot originate from more than one source. Also, the entries in the columns corresponding to hypotheses H_{FT} and H_{NT} may sum to a number greater than one since any, all, or none of the measurements may have originated from false sources or potential new targets; the entries in the columns corresponding to H_{DT} may only sum to zero or one, since the actual target (pre-existing track) is assumed to generate at most one measurement.

The remainder of this section is dedicated to finding a rigorous Bayesian solution for the measurement origin uncertainty problem encountered in target tracking. The measurement-oriented data association method is chosen over the target-oriented data association method because it is the more general of the two methods, and the target-oriented method is readily obtained from the measurement-oriented approach. Because measurement origin uncertainty is modeled in the same way as the kinematics model parameter uncertainty for switching models, the recursive Bayesian solution is mathematically similar to the solution found in Subsection 2.4.2. Consequently, the recursive Bayesian solution suffers from the same computational difficulties as the switching model solution. Specifically, the number of association events, or hypotheses, grows exponentially over time, and some type of hypothesis reduction routine is necessary if one wishes to implement this solution in a practical tracking system.

Approximations of the recursive Bayesian solution for the measurement uncertainty problem may be divided into two categories¹⁴. The first classification includes the Probabilistic Data Association Filter (PDAF) (for single target scenarios), Joint Probabilistic Data Association Filter (JPDAF) and its variants (for multiple target scenarios), and N-Scan filter with N set to one [5,32]. PDAF and JPDAF approximate

¹⁴Measurement gating is another type of approximation to the rigorous Bayesian solution. This approximation reduces the number of terms necessary for computing the target state pdf by effectively limiting the sample space of Θ_{i_k} through the restrictions imposed by the measurement gates.

the rigorous solution in the same manner as the GPB-1 approximation in which the target state Gaussian mixture pdf is approximated by a *single* Gaussian pdf at the end of each scan cycle [5]. The N-scan filter, with N set to one, uses an approximation similar to that used for the GPB-2 algorithm [5]. These three algorithms rely on the target-oriented data association method, so the number of targets is assumed known a priori. The second category includes the Joining and Clustering [30, 31] and Integral Square Error [38, 40, 41] mixture reduction algorithms, which are approximations that may use either of the two data association methods. Both algorithms reduce the number of mixture components in the target state Gaussian mixture pdf at the end of each scan based on preset criteria and, in most cases, provide a better approximation to the target state pdf than a single Gaussian pdf approximation.

2.5.1 A Bayesian Solution for Measurement Origin Uncertainty. Although this thesis only considers tracking a single target in clutter using the target-oriented data association method, a Bayesian solution for tracking multiple targets in the presence of measurement origin uncertainty utilizing the measurement-oriented data association method is developed in this subsection based on [5, 27, 38]. The reason multiple targets and the measurement-oriented data association method are considered is two-fold. First, the multiple-target solution using the measurement-oriented data association method is more general. Second, this solution readily reduces to the single-target, target-oriented data association solution.

A new set of notation is needed for the multiple target tracking in the presence of measurement origin uncertainty scenario. When considering more than one target, the state random process vector must include the state random process vectors of all of the targets. Thus, a joint target state random process composite vector is formed as

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_{N_T(k)}(k) \end{bmatrix} \quad (2.60)$$

where $N_T(k)$ is the number of targets at sample k . The number of targets is constrained to equal the sum of the number of detected targets, N_{DT} , and the number of new targets, N_{NT} , at scan k . Since potentially more than one measurement vector is available at each scan due to measurement origin uncertainty, the composite measurement *random process* vector is represented by

$$\mathbf{Z}(k) = \begin{bmatrix} \mathbf{z}_1(k) \\ \vdots \\ \mathbf{z}_{N_m(k)}(k) \end{bmatrix} \quad (2.61)$$

where $N_m(k)$ is the *random* number of measurements at sample k . Upon observation of the measurements, the composite measurement random process vector becomes the *realized* composite measurement vector given by

$$\bar{\mathbf{Z}}_k = \begin{bmatrix} \mathbf{z}_{k,1} \\ \vdots \\ \mathbf{z}_{k,N_m(k)} \end{bmatrix}. \quad (2.62)$$

Using the new definition of the composite measurement vector, the measurement history composite vector of Subsection 2.2.1 becomes

$$\mathbf{Z}^k = \begin{bmatrix} \mathbf{Z}(1) \\ \vdots \\ \mathbf{Z}(k) \end{bmatrix}. \quad (2.63)$$

If one only considers a single-target scenario utilizing the target-oriented data association method, then Equation (2.60) is simply a single target state vector and the number of new targets, N_{NT} , is zero. Equations (2.61), (2.62), and (2.63) remain unchanged. Using this new notation, a recursive Bayesian solution may be formulated in the same manner as in Subsection 2.4.2.

The joint target state random process composite vector conditioned on the measurement history is

$$\begin{aligned}
f(\mathbf{X}(k), \Theta(k) | \mathbf{Z}^k) &= f(\mathbf{X}(k) | \Theta(k), \mathbf{Z}^k) f(\Theta(k) | \mathbf{Z}^k) \\
&= f(\mathbf{X}(k) | \Theta(k), \mathbf{Z}^k) \sum_{i_k \in N_H(k)} p(\Theta(k) = \Theta_{i_k} | \mathbf{Z}^k) \delta(\Theta(k) - \Theta_{i_k}) \\
&= \sum_{i_k \in N_H(k)} f(\mathbf{X}(k) | \Theta_{i_k}, \mathbf{Z}^k) p(\Theta_{i_k} | \mathbf{Z}^k). \tag{2.64}
\end{aligned}$$

The second line of this equation follows from the generalized definition of a pdf (see Subsection 2.4.1) since $\Theta(k)$ is a discrete random process vector, and $N_H(k)$ represents the number of hypotheses or association events in the discrete sample space of the association event discrete random process vector Θ_{i_k} (see Figure 2.10 for an example sample space). A *deferred decision* approach is desired so that the previous association events are incorporated into the decision criterion for evaluating association hypotheses. Therefore, the Bayesian solution should include the entire *association event history*, $\{\Theta_{i_\ell}\}_1^k$. The above equation is then written in terms of the joint pdf of $\Theta(k)$ for all time as

$$f(\mathbf{X}(k), \Theta(k) | \mathbf{Z}^k) = \sum_{i_k \in N_H(k)} \cdots \sum_{i_1 \in N_H(1)} f(\mathbf{X}(k) | \{\Theta_{i_\ell}\}_1^k, \mathbf{Z}^k) p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k). \tag{2.65}$$

As in the switching model case, the Bayesian solution results in a Gaussian mixture (in this case, mixtures of mixtures). It is computationally intractable unless approximated using one of the methods presented in the section introduction. The remainder of this subsection is dedicated to determining $p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k)$, since $f(\mathbf{X}(k) | \{\Theta_{i_\ell}\}_1^k, \mathbf{Z}^k)$ is provided by the Kalman filter solution given that there is no measurement origin uncertainty.

The joint conditional pmf, $p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k)$, of the current association event history, given the measurement history, \mathbf{Z}^k , is [5]

$$\begin{aligned} p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k) &= \frac{f(\Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}(k), \mathbf{Z}^{k-1}, N_m(k))}{f(\mathbf{Z}(k), \mathbf{Z}^{k-1}, N_m(k))} \\ &= \frac{f(\mathbf{Z}(k) | \Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) p(\Theta_{i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) p(\{\Theta_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1})}{f(\mathbf{Z}(k) | \mathbf{Z}^{k-1}, N_m(k))} \end{aligned} \quad (2.66)$$

for $i_k = 1, \dots, N_H(k)$, $i_\ell = 1, \dots, N_H(\ell)$, and $\ell = 1, \dots, k-1$. Notice the conditioning on the number of measurements at sample k^{15} , $N_m(k)$, is included in the above pdfs/pdfs since this quantity determines the number of hypotheses in the sample space of Θ_{i_k} (see the example in Figure 2.10, for instance). Furthermore, conditioning on the current measurement history \mathbf{Z}^k implies conditioning on the number of measurements, $N_m(k)$, even though this conditioning is not explicitly shown in the left-hand side of Equation (2.66). Also, the conditioning on $N_m(k)$ was dropped in the pmf $p(\{\Theta_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1})$ because the number of current measurements has no bearing on the prior association event history. Equation (2.66) is evaluated in the next three subsections.

¹⁵This number represents the *total number* of measurements made in the *entire* surveillance region so that a new track initiation capability is maintained [27].

2.5.1.1 The Composite Measurement Likelihood Function.

The first term in the numerator of the second line of Equation (2.66), the composite measurement likelihood function, is evaluated in this subsection. The distribution of each measurement random process vector contained in the composite measurement vector in Equation (2.61) is assumed to be either Gaussian, for measurements hypothesized originating from existing targets in a previous scan, or Uniform, for measurements hypothesized originating from new targets or false sources¹⁶. For each realized measurement vector at sample k , $\mathbf{z}_{k,j}$, hypothesized under association event Θ_{i_k} ($i_k = 1, \dots, N_H(k)$) as originating from a corresponding existing target j ($j = 1, \dots, N_{DT,i_k}$) that was hypothesized in a previous scan and detected in the current scan, the measurement likelihood function for the j^{th} existing target is

$$f(\mathbf{z}_j(k)|\Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k))|_{\mathbf{z}_j(k)=\mathbf{z}_{k,j}} = L(\mathbf{z}_{k,j}; \mathbf{r}_{k,j}, \mathbf{S}_j(k)) \quad (2.67)$$

$$= \frac{\exp\left[-\frac{1}{2}\mathbf{r}_{k,j}^T \mathbf{S}_j^{-1}(k) \mathbf{r}_{k,j}\right]}{(2\pi)^{\frac{m}{2}} \sqrt{\det \mathbf{S}_j(k)}}$$

$$\mathbf{r}_{k,j} = \mathbf{z}_{k,j} - \mathbf{H}_j \hat{\mathbf{x}}_j(k|k-1)$$

$$\mathbf{S}_j(k) = \mathbf{H}_j \mathbf{P}_j(k|k-1) \mathbf{H}_j^T + \mathbf{R}_j(k).$$

The measurement likelihood function for a measurement hypothesized under association event Θ_{i_k} as originating from a potential new target or false source is represented by $1/V_S$, where V_S is the surveillance volume [5, 27].

¹⁶In previous sections of this chapter, a measurement originating from a target is assumed to be distributed as a Gaussian random quantity since the target state random process vector is assumed to be Gaussian. In this section, a measurement could also originate from a new target or a false source. Measurements originating from new targets or false sources are assumed to appear at random locations in the surveillance volume with equal probability. Thus, a measurement hypothesized to originate from a new target or false source is mathematically modeled as a uniformly-distributed random variable.

Typically, the target state random process vectors, which form the joint target state random process composite vector in Equation (2.60), are assumed to be mutually independent [5]. Using this assumption and the assumption that the measurement noise process vector in Equation (2.4) is independent of the state process vector, the measurement random process vectors forming the composite measurement vector in Equation (2.61) are also mutually independent¹⁷. To evaluate the composite likelihood function, $f(\mathbf{Z}(k)|\Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k))$, first note that the total number of measurements at scan k , $N_m(k)$, is equal to the sum of the number of measurements hypothesized under a given association event, Θ_{i_k} , as originating from targets that were hypothesized in a previous scan and detected in the current scan (N_{DT,i_k}), potential new targets (N_{NT,i_k}), and false sources (N_{FT,i_k}). Invoking the mutual independence of the measurement random process vectors and the modeling assumptions in the previous paragraph, the composite measurement likelihood function may be written as [5, 27]

$$\begin{aligned}
& f(\mathbf{Z}(k)|\Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k))|_{\mathbf{Z}(k)=\mathbf{z}_k} \\
&= \left(\frac{1}{V_S} \right)^{N_{FT,i_k} + N_{NT,i_k}} \cdot L \left(\{\mathbf{z}_{k,j}\}_1^{N_{DT,i_k}}; \{\mathbf{r}_{k,j}\}_1^{N_{DT,i_k}}, \{\mathbf{S}_j(k)\}_1^{N_{DT,i_k}} \right) \\
&= \left(\frac{1}{V_S} \right)^{N_{FT,i_k} + N_{NT,i_k}} \cdot \prod_{j=1}^{N_{DT,i_k}} \frac{\exp \left[-\frac{1}{2} \mathbf{r}_{k,j}^T \mathbf{S}_j^{-1}(k) \mathbf{r}_{k,j} \right]}{(2\pi)^{\frac{m}{2}} \sqrt{\det \mathbf{S}_j(k)}} \tag{2.68}
\end{aligned}$$

for $i_k = 1, \dots, N_H(k)$. The terms $\mathbf{r}_{k,j}$ and \mathbf{S}_j are defined in Equation (2.67). If measurement gating is used, then measurements falling outside of the gate of an existing target cannot be hypothesized to have originated from that target. In addition, if an association event consists of hypotheses which do not associate any measurement to

¹⁷One possible intuitive justification for this assumption lies in the previously-made condition that a measurement cannot originate from multiple sources. Since each measurement is assumed to originate from a distinct source, it seems intuitive to believe that the measurements are mutually independent.

existing targets hypothesized in the previous scan and detected in the current scan, then N_{DT,i_k} is zero for that association event, and the product term, $\prod_{j=1}^0(\cdot)$, in Equation (2.68) is not evaluated.

Equation (2.68) may be applied to the example given in the section introduction and shown in Figure 2.10. Consider association event Θ_{7_k} in which both measurements are hypothesized to have originated from false sources. For this association event, the composite measurement likelihood becomes $(1/V_S)^2$ since $N_{DT,7_k}$ and $N_{NT,7_k}$ equal zero, and $N_{FT,7_k}$ is two (recall that hypothesis H_{FT} supposes that a measurement is due to a false source). As another example of using Equation (2.68), consider association event Θ_{2_k} (in the same figure) which hypothesizes that the first measurement is due to the existing target hypothesized in a previous scan (as indicated by hypothesis H_{DT}), and that the other measurement is due to a new target (hypothesized under H_{NT}). Its composite measurement likelihood function is $(1/V_S)L(\mathbf{z}_{k,1}; \mathbf{r}_{k,1}, \mathbf{S}_1(k))$.

2.5.1.2 The Conditional Current Association Event pmf. Consider the conditional current association event pmf, conditioned on the current number of measurements and the prior association event and measurement histories, $p(\Theta_{i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k))$. To evaluate this pmf, one must either wisely choose a discrete probability distribution model for Θ_{i_k} (e.g., discrete Uniform, Poisson, Multinomial, etc.) or use some other assumption about the random quantity Θ_{i_k} . Any choice of probability distribution model that does not incorporate prior knowledge about the detection capabilities of the tracking system or other engineering insights, such as the number of expected false-origin measurements or potential new targets in a given surveillance volume, would be unwise. Such prior knowledge could be incorporated into the current association event pmf through N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} by modeling these integer-valued variables as appropriately-chosen random variables. This approach is taken in [5, 27], and it will be reproduced in this subsection.

First, model the number of measurements hypothesized according to the i_k^{th} association event as originating from existing targets hypothesized in a previous scan

and detected in the current scan (N_{DT,i_k}), potential new targets (N_{NT,i_k}), and false sources (N_{FT,i_k}) as integer-valued random variables. Next, form the *joint* conditional pmf of the random quantities Θ_{i_k} , N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} . Applying the conditional probability rule for pmfs to this joint conditional pmf yields the desired result

$$\begin{aligned} p(\Theta_{i_k}, N_{DT,i_k}, N_{NT,i_k}, N_{FT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ = p(\Theta_{i_k} | N_{DT,i_k}, N_{NT,i_k}, N_{FT,i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ \cdot p(N_{DT,i_k}, N_{NT,i_k}, N_{FT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)). \end{aligned} \quad (2.69)$$

The conditional pmf of the current association event is now conditioned on the information one may possess about the detection capabilities, the expected number of false-origin measurements, and the expected number of potential new targets. This conditioning achieves the desired goal of the previous paragraph. The second conditional pmf in Equation (2.69) depends on the probability distribution assigned to each of the integer-valued random variables N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} . Both pmfs on the right-hand side of Equation (2.69) are evaluated next.

If all association events containing the same number of detected targets, the same number of new target measurements, and the same number of false-origin measurements are considered equally likely [38], then the new conditional current association event pmf $p(\Theta_{i_k} | N_{DT,i_k}, N_{NT,i_k}, N_{FT,i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k))$ may be evaluated by counting methods, like those found in [20]. Assuming that measurement gating is not used (to simplify the matter¹⁸), the total number of association events is found by calculating the product of the number of ways to partition $N_m(k)$ measurements into N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} mutually exclusive and exhaustive partitions, and the number of ways to assign N_{DT,i_k} measurements to the existing targets hypothesized

¹⁸Measurement gating imposes a restriction on associations between observed measurements and existing targets hypothesized in a previous scan since observations outside the measurement gate of an existing target may not be assigned to that target. This restriction makes developing a general equation for the total number of association events very difficult due to the random nature of the measurements (i.e., one does not know a priori where measurements may fall within a surveillance volume).

under the association event history through sample $k - 1$, $\{\Theta_{i_\ell}\}_1^{k-1}$. The first quantity in the product is found by using the multinomial coefficient [20, 27],

$$\frac{N_m(k)!}{N_{DT,i_k}! N_{NT,i_k}! N_{FT,i_k}!}, \quad (2.70)$$

since this coefficient is the number of ways to partition the total number of measurements at scan k into three mutually exclusive and exhaustive groups. The number of ways to assign the hypothesized number of measurements associated with detected targets in the current scan k , N_{DT,i_k} , to the total number of existing targets hypothesized under the association event history through sample $k - 1$, denoted as N_{TGT} , is [27]

$$\frac{N_{TGT}!}{(N_{TGT} - N_{DT,i_k})!}. \quad (2.71)$$

(Notice that Equation (2.71) is the “sampling without replacement and with ordering” equation found in [20].) Combining Equations (2.70) and (2.71) yields the new conditional current association event pmf (the first pmf on the right-hand side of Equation (2.69)) [27]:

$$\begin{aligned} p(\Theta_{i_k} | N_{DT,i_k}, N_{NT,i_k}, N_{FT,i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ = \left[\frac{N_m(k)!}{N_{DT,i_k}! N_{NT,i_k}! N_{FT,i_k}!} \cdot \frac{N_{TGT}!}{(N_{TGT} - N_{DT,i_k})!} \right]^{-1}. \end{aligned} \quad (2.72)$$

Technically, Equation (2.72) is not a true pmf since it is not normalized. A normalization factor could be included in this expression, however, since the joint conditional pmf of the current association event history, Equation (2.66), is normalized, normalization of Equation (2.72) is omitted.

The joint conditional pmf of N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} , which appears in the last line of Equation (2.69), may be found by first assuming that the random variables

N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} are mutually independent [5]¹⁹. Then, the joint pmf of these random variables becomes

$$\begin{aligned} p(N_{DT,i_k}, N_{NT,i_k}, N_{FT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ = p(N_{DT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) p(N_{NT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ \cdot p(N_{FT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)). \end{aligned} \quad (2.73)$$

The random variable N_{DT,i_k} is modeled as Binomial with parameters P_D , the probability of detection, and N_{TGT} , the total number of existing targets hypothesized under the association event history through sample $k-1$. The pmf of this Binomial random variable is [20, 27]

$$\begin{aligned} p(N_{DT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ = \frac{N_{TGT}!}{N_{DT,i_k}!(N_{TGT} - N_{DT,i_k})!} P_D^{N_{DT,i_k}} (1 - P_D)^{N_{TGT} - N_{DT,i_k}}. \end{aligned} \quad (2.74)$$

Poisson random variables are used to model N_{NT,i_k} and N_{FT,i_k} , and the pmf for each random variable is [20, 27]

$$\begin{aligned} p(N_{NT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) &= \frac{(\lambda_{NT} V_S)^{N_{NT,i_k}}}{N_{NT,i_k}!} e^{-\lambda_{NT} V_S} \\ p(N_{FT,i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) &= \frac{(\lambda_{FT} V_S)^{N_{FT,i_k}}}{N_{FT,i_k}!} e^{-\lambda_{FT} V_S}. \end{aligned} \quad (2.75)$$

Since the parameter of a Poisson pmf has the units of “average number of events,” the constants $\lambda_{NT} V_S$ and $\lambda_{FT} V_S$ may be interpreted as the expected number of new targets and false-origin measurements in a given surveillance volume, V_S . The terms λ_{FT} and

¹⁹Williams [38] pointed out that the conditioning on $N_m(k)$ in this pmf makes this independence assumption questionable since the random variables N_{DT,i_k} , N_{NT,i_k} , and N_{FT,i_k} are mathematically related by the equation $N_m(k) = N_{DT,i_k} + N_{NT,i_k} + N_{FT,i_k}$. However, it was shown in [38] that the mutual independence assumption may be made without conditioning on the number of current measurements by applying Bayes’ rule to remove this conditioning.

λ_{NT} are the false-origin measurement clutter density and new target measurement clutter density, respectively.

Finally, combining Equations (2.72), (2.74), and (2.75), and canceling the terms $N_{TGT}!$, $N_{DT,i_k}!$, $(N_{TGT} - N_{DT,i_k})!$, $N_{FT,i_k}!$, and $N_{NT,i_k}!$, the conditional current association event pmf under the measurement-oriented data association assumption and in the absence of measurement gating is [27]

$$p(\Theta_{i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) = \frac{1}{N_m(k)!} \cdot P_D^{N_{DT,i_k}} (1 - P_D)^{N_{TGT} - N_{DT,i_k}} \quad (2.76)$$

$$\cdot (\lambda_{FT} V_S)^{N_{FT,i_k}} e^{-\lambda_{FT} V_S} \cdot (\lambda_{NT} V_S)^{N_{NT,i_k}} e^{-\lambda_{NT} V_S}$$

for $i_k = 1, \dots, N_H(k)$. Notice that this pmf includes information a designer might possess about the detection capabilities of a tracking system and the expected number of false-origin and new target measurements generated by the tracker in a certain surveillance volume. Also, this pmf may be modified in the case of a target-oriented data association approach by setting N_{NT,i_k} to zero with probability one.

2.5.1.3 The Normalization Factor. The final piece of Equation (2.66), the normalization factor in the denominator, is evaluated in this subsection. Since the joint pdf of the entire measurement history, the current number of measurements, and the entire association event history is assumed to exist (as shown in the first line of Equation (2.66)), the normalization factor may be expanded using marginal probabilities as

$$f(\mathbf{Z}(k) | \mathbf{Z}^{k-1}, N_m(k)) = \sum_{i_k \in N_H(k)} \cdots \sum_{i_1 \in N_H(1)} f(\mathbf{Z}(k), \Theta_{i_k}, \dots, \Theta_{i_1} | \mathbf{Z}^{k-1}, N_m(k)).$$

Two applications of the law of conditional probability for pdfs/pmf yields the desired result

$$\begin{aligned} f(\mathbf{Z}(k) | \mathbf{Z}^{k-1}, N_m(k)) &= \sum_{i_k \in N_H(k)} \cdots \sum_{i_1 \in N_H(1)} f(\mathbf{Z}(k) | \Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) \\ &\quad \cdot p(\Theta_{i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) p(\{\Theta_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1}). \end{aligned} \quad (2.77)$$

Finally, recall Equation (2.66), the joint conditional pmf of the current association event history:

$$\begin{aligned} p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k) &= \frac{f(\Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}(k), \mathbf{Z}^{k-1}, N_m(k))}{f(\mathbf{Z}(k), \mathbf{Z}^{k-1}, N_m(k))} \\ &= \frac{f(\mathbf{Z}(k) | \Theta_{i_k}, \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) p(\Theta_{i_k} | \{\Theta_{i_\ell}\}_1^{k-1}, \mathbf{Z}^{k-1}, N_m(k)) p(\{\Theta_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1})}{f(\mathbf{Z}(k) | \mathbf{Z}^{k-1}, N_m(k))} \end{aligned} \quad (2.66)$$

Substituting Equations (2.68), (2.76), and (2.77) into this expression and canceling the term $(1/V_S)^{N_{FT,i_k} + N_{NT,i_k}}$ with the term $V_S^{N_{FT,i_k} + N_{NT,i_k}}$ in both the numerator and denominator, and canceling the terms $e^{-\lambda_{FT}V_S}$ and $e^{-\lambda_{NT}V_S}$ yields the intermediate

result

$$p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k) = \frac{NUM}{DEN} \quad (2.78)$$

$$NUM = \prod_{j=1}^{N_{DT,i_k}} \frac{\exp\left[-\frac{1}{2}\mathbf{r}_{k,j}^T \mathbf{S}_j^{-1}(k) \mathbf{r}_{k,j}\right]}{(2\pi)^{\frac{m}{2}} \sqrt{\det \mathbf{S}_j(k)}} P_D^{N_{DT,i_k}} (1 - P_D)^{N_{TGT} - N_{DT,i_k}}$$

$$\cdot (\lambda_{FT} V_S)^{N_{FT,i_k}} (\lambda_{NT} V_S)^{N_{NT,i_k}} p(\{\Theta_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1})$$

$$DEN = \sum_{i_k \in N_H(k)} \cdots \sum_{i_1 \in N_H(1)} \prod_{j=1}^{N_{DT,i_k}} \frac{\exp\left[-\frac{1}{2}\mathbf{r}_{k,j}^T \mathbf{S}_j^{-1}(k) \mathbf{r}_{k,j}\right]}{(2\pi)^{\frac{m}{2}} \sqrt{\det \mathbf{S}_j(k)}} p(\{\Theta_{i_\ell}\}_1^{k-1} | \mathbf{Z}^{k-1}) \\ \cdot P_D^{N_{DT,i_k}} (1 - P_D)^{N_{TGT} - N_{DT,i_k}} (\lambda_{FT} V_S)^{N_{FT,i_k}} (\lambda_{NT} V_S)^{N_{NT,i_k}}.$$

The final result is obtained by substituting Equation (2.66) into Equation (2.65),

$$f(\mathbf{X}(k), \Theta(k) | \mathbf{Z}^k) = \sum_{i_k \in N_H(k)} \cdots \sum_{i_1 \in N_H(1)} p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k) f(\mathbf{X}(k) | \{\Theta_{i_\ell}\}_1^k, \mathbf{Z}^k) \quad (2.65)$$

which is a multivariate Gaussian mixture with $N_H(k) \cdot N_H(k-1) \cdots N_H(2) \cdot N_H(1)$ components. At each new scan, another summation is added to this equation, and the Bayesian solution cannot be implemented without approximation.

Approximating Equation (2.65) to trim computations while maintaining good performance is the focus of this thesis. Effectively, the approximation will reduce the original $N_H(k) \cdot N_H(k-1) \cdots N_H(2) \cdot N_H(1)$ Gaussian mixture components in Equation (2.65) to some manageable level. In practice, this reduction will be accomplished by a *mixture reduction algorithm* (MRA) which will reduce the $N_H(k)$ number of association events (or, equivalently, mixture components) to $N_R(k)$ association events at the end of each scan k .

2.5.2 Tracking with Measurement Origin Uncertainty Summary. Measurement origin uncertainty leads to the data association problem in which the source of a measurement is ambiguous. One poses hypotheses about the possible source of each observed measurement as a first-step towards solving this problem. Hypotheses are formed according to one of two assumptions about the potential origin of a measurement used in practice. The target-oriented data association approach assumes that the potential sources of measurements are existing tracks that were hypothesized in previous scans or false sources, while the measurement-oriented data association method supposes that measurements may also arise from potential new tracks in addition to existing tracks and false sources. Both data association approaches restrict the number of measurements generated by any single source to one.

A rigorous Bayesian solution for tracking multiple targets in the presence of measurement origin uncertainty using the measurement-oriented data association approach was presented in this section as a second step towards solving the data association problem. Association events were modeled as a discrete random process vector, $\Theta(k) = \Theta_{i_k}$, and were included with the joint target state random process composite vector $\mathbf{X}(k)$ as another random quantity to be estimated. The resulting Bayesian solution was mathematically similar to that for the switching model case of Subsection 2.4.2 since the measurement origin uncertainty was modeled in the same manner as the kinematics model parameter uncertainty for switching models. Information about the detection capabilities of a tracking system and the expected number of false-origin and new target measurements generated by the tracker in a certain surveillance volume were embedded into the pmf of the association event history. Where appropriate, the steps necessary to convert the multiple-target, measurement-oriented data association method Bayesian solution into the single-target, target-oriented data association approach solution were noted.

The final form of the solution is a Gaussian mixture with $N_H(k) \cdots N_H(1)$ components (as shown in Equation (2.65)). This solution is intractable, and some type of approximation is necessary to make the solution a viable candidate for practical im-

plementation. Two fundamental approximations are to reduce the Gaussian mixture to a single component, which is used by PDAF and JPDAF, or to a lower number of components, which is the approximation used by the Joining and Clustering and the Integral Square Error cost-function-based MRAs, at the end of each scan k . The focus of this thesis is on the latter kind of approximation, specifically the type of MRA introduced by Williams in [38, 40, 41].

2.6 Summary

This chapter introduced target tracking as a means of determining the state of targets over some time interval of interest from observations of the targets in the presence of uncertainty. Two basic sources of uncertainty are due to mathematical models of the targets' dynamics which may, at best, only approximate the true motion of the targets and may change substantially from one time instant to another, and from sensor noise which corrupts measurements. If the target kinematics model and measurement model are linear, and all random quantities are modeled as Gaussian, then the Kalman filter, which is a linear recursive Bayesian filter, provides the optimal mean and covariance estimates of the target state random process vector under almost all practical criteria, conditioned on an assumed measurement association history and an assumed target dynamics model. If either model is nonlinear, then a nonlinear recursive Bayesian filter is used instead. However, the nonlinear filter will, in general, not produce an optimal estimate of the target state process vector.

Gaussian mixtures result from using Bayes estimation to solve target tracking problems in which kinematics model parameter and measurement origin uncertainty exists. The general form of a multivariate Gaussian mixture pdf was presented, and equations for calculating the overall mean and covariance of a target state random process vector described by a multivariate Gaussian mixture pdf were provided. Also, the effects of merging or deleting mixture components on the resulting Gaussian mixture during a mixture reduction process were described.

Bayesian solutions for the target state pdf in the presence of kinematics model parameter²⁰ and measurement origin uncertainty were also derived using Bayes estimation. Solutions which modeled the uncertainty as a random vector were tractable, while Bayesian solutions which represented the uncertainty as a random process vector were intractable, and approximation was necessary to implement the solutions. In fact, new methods for approximating the rigorous Bayesian solution for the target state Gaussian mixture pdf in the presence of measurement origin uncertainty is the focus of this thesis, and such methods will be presented in subsequent chapters.

²⁰Kinematics model parameter uncertainty should not be confused with uncertainty in a kinematics model due to mathematically modeling a target's dynamics. The former source of uncertainty arises from a designer not knowing which model to use, while the latter source of uncertainty is a simple admission that mathematical equations cannot exactly describe the realistic motion of an object.

III. Estimating Probability Density Functions

In [38], Williams derived the Maximum Likelihood measure which he believes “... is probably the most physically meaningful cost function for this application [Gaussian mixture reduction].” Because one aspect of the goal for this thesis is to develop a new mixture reduction algorithm which outperforms any previously published algorithm, Williams’ endorsement of the Maximum Likelihood measure is the motivation behind this chapter. As such, Chapter III explores the techniques of pdf estimation drawn from statistical inference which is based on the well-developed field of the mathematical theory of probability and mathematical statistics (to include maximum likelihood estimation). Given a set of random observations, one tries to “infer” the underlying distribution that spawned these samples. In some cases this distribution is known, or at least assumed known, to be of a certain type (Gaussian, Poisson, etc.) and the task is to estimate the parameter or parameters of the distribution (e.g., the mean and variance for a Gaussian density, the rate for a Poisson distribution, etc.). In other cases the distribution may be known (or assumed known) to be limited to some set of possible distributions, and the task now is to identify the correct one from the set and estimate the parameter or parameters of this distribution.

Although pdf estimation may not seem directly applicable to the purpose of approximating a Gaussian mixture with one containing a lower number of components, it is useful to explore the concepts and techniques of this field in the hope of gaining insights into an appropriate method for mixture approximation (such as the Maximum Likelihood measure which Williams developed in [38]). Generally, there are at least two methods of estimation that can be applied to this problem: maximum likelihood estimation (MLE) and Bayesian estimation. Both methods attempt to estimate the one or many parameters of a presumed pdf, but differ in application. MLE is used when the pdf parameters are deterministic (fixed but unknown) while Bayesian estimation can handle the case in which the parameters are random. In this thesis, the parameters of a Gaussian mixture will be modeled as deterministic quan-

ties, and this chapter uses MLE to solve for these parameters¹. This chapter also introduces an iterative implementation of MLE called the Expectation Maximization (EM) algorithm.

The following sections are not intended to be a comprehensive narrative of MLE methods, but rather an introduction highlighting certain aspects of this approach that may shed light on approximating a full-component Gaussian mixture with one having a lower number of mixture components. Complete treatments of MLE are presented in [10, 17, 25, 42], from a mathematical statistics perspective, and [21, 22, 35, 36], from an engineering viewpoint.

3.1 Maximum Likelihood Estimation

The method of maximum likelihood can be traced back to Gauss but was not applied to general estimation problems until R. A. Fisher published a short paper on the topic in 1912 [10]. Over the next thirty years Cramér [10], Rao [25], and others developed this method in a more formal mathematical manner. As their work popularized the method of maximum likelihood, it became commonly known as maximum likelihood estimation (MLE) and as the preferred way to estimate the deterministic parameters of a pdf.

MLE is ideally suited to estimating one or more deterministic parameters of a pdf when independent samples are drawn from a known distribution. That is, given a set of independent identically distributed (i.i.d.) observations from a random quantity for which the mathematical form of the pdf is known (except for a certain number of parameters), the MLE approach may be used to find an estimate of the pdf parameters. This estimate is a random quantity since it is a function of the random observations, and under the Cramér-Rao (C.R.) regularity conditions (see [42]

¹In Chapter II, Bayesian estimation was used to estimate the target state since it was modeled as a *random process* vector. However, this chapter does not consider Bayesian estimation for the Gaussian mixture parameters since they are modeled as unknown *deterministic* quantities.

pp. 182–183), the estimate has the asymptotic (the number of observations is large) qualities of [10, 25, 42]:

- converging to the true value of the parameters,
- converging to a Gaussian distribution, and
- efficiency (the distribution of the estimate is minimum variance).

A detailed derivation of these properties is contained in Subsection 3.1.1 to help the reader better understand the asymptotic qualities of MLE. For instance, the derivation will show that the first property listed above is not always true since the maximization may only converge to one of a number of local maxima.

As an example, consider the problem of estimating a single parameter of a pdf $f(z|\alpha)$ when a set of n i.i.d. observations $\{z_i\}_1^n$ are made. The *likelihood function* is defined as

$$L(\{z_i\}_1^n; \alpha) = \prod_{i=1}^n f(z_i|\alpha) \quad (3.1)$$

since the observations are i.i.d (i.e., they are independent so the joint density of the n observations is just the product of the separate marginal densities and the marginal density has the same form). The method of maximum likelihood is simply maximizing this expression for some α in the open interval A (α is not an endpoint). In theory the estimate from this maximization will converge to the true value of the parameter as the number of observations grows; however, there is no guarantee that the maximization is global, and the true value of the parameter may not be so easily found. This fact will become evident in the proof that follows.

3.1.1 Asymptotic Properties of MLE. The three asymptotic qualities of MLE will be proven for a single parameter in a derivation according to [10, 25, 42]. But before continuing, it is necessary to introduce some preliminary information. First, the product form of $L(\cdot)$ is converted into a summation so that useful convergence theorems may be invoked. Since $\ln \prod_{i=1}^n y_i = \sum_{i=1}^n \ln(y_i)$ and $\ln(\cdot)$ is a monotonically increasing function, the value of α that maximizes Equation (3.1) also maximizes the

log-likelihood function

$$\ln L(\{z_i\}_1^n; \alpha) = \sum_{i=1}^n \ln f(z_i|\alpha). \quad (3.2)$$

By solving the *likelihood equation*

$$\frac{\partial \ln L(\{z_i\}_1^n; \alpha)}{\partial \alpha} \Big|_{\alpha=\hat{\alpha}_{ml}} = 0 \quad (3.3)$$

for $\hat{\alpha}_{ml}$ one obtains the *maximum likelihood estimate* of the parameter α . Second, the following conditions are imposed [10, 25, 42]:

- (i) $\alpha \in A \subset \mathbb{R}$, α not an endpoint of A .
- (ii) For almost all z , $\ln f(z|\alpha)$ is analytic² in α .
- (iii) The observations, z_i , are i.i.d..
- (iv) For every $\alpha \in A$,

$$\int_{-\infty}^{\infty} \left(\frac{\partial \ln f(z|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}^2 f(z|\alpha_o) dz = k^2 < \infty \quad (3.4)$$

where α_o is the *true* value of the parameter and $k > 0$.

- (v) All moments of $\frac{\partial \ln f(z|\alpha)}{\partial \alpha}$ are finite; i.e.

$$E \left\{ \frac{\partial^i \ln f(z|\alpha)}{\partial \alpha^i} \right\} < \infty. \quad (3.5)$$

Third, the following identities will be used.

- (a) By the chain rule of calculus,

$$\frac{\partial \ln f(z|\alpha)}{\partial \alpha} = \frac{1}{f(z|\alpha)} \frac{\partial f(z|\alpha)}{\partial \alpha}.$$

²An analytic function is guaranteed an infinite number of finite-valued derivatives [3].

(b) The first and second partial derivatives of $\int_{-\infty}^{\infty} f(z|\alpha_o) dz$ are zero since

$$\int_{-\infty}^{\infty} (f(z|\alpha))_{\alpha=\alpha_o} dz = 1$$

and differentiating this expression with respect to α yields:

$$\frac{d}{d\alpha} \int_{-\infty}^{\infty} (f(z|\alpha))_{\alpha=\alpha_o} dz = \frac{d}{d\alpha} 1.$$

Thus:

$$\int_{-\infty}^{\infty} \left(\frac{\partial f(z|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o} dz = 0 \quad (3.6)$$

$$\int_{-\infty}^{\infty} \left(\frac{\partial^2 f(z|\alpha)}{\partial \alpha^2} \right)_{\alpha=\alpha_o} dz = 0. \quad (3.7)$$

(c) $E\{[(\partial^2/\partial\alpha^2) \ln f(z|\alpha)]_{\alpha=\alpha_o}\} = -E\{[(\partial/\partial\alpha) \ln f(z|\alpha)]_{\alpha=\alpha_o}^2\}$ since

$$\begin{aligned} E \left\{ \left(\frac{\partial^2 \ln f(z|\alpha)}{\partial \alpha^2} \right)_{\alpha=\alpha_o} \right\} &= E \left\{ \frac{\partial}{\partial \alpha} \left(\frac{\frac{\partial f(z|\alpha)}{\partial \alpha}}{f(z|\alpha)} \right)_{\alpha=\alpha_o} \right\} \\ &= E \left\{ \left(\frac{\frac{\partial^2 f(z|\alpha)}{\partial \alpha^2} f(z|\alpha) - \frac{\partial f(z|\alpha)}{\partial \alpha} \frac{\partial f(z|\alpha)}{\partial \alpha}}{f^2(z|\alpha)} \right)_{\alpha=\alpha_o} \right\} \\ &= E \left\{ \left(\frac{\frac{\partial^2 f(z|\alpha)}{\partial \alpha^2}}{f(z|\alpha)} \right)_{\alpha=\alpha_o} \right\} - E \left\{ \left(\frac{\frac{\partial f(z|\alpha)}{\partial \alpha}}{f(z|\alpha)} \right)_{\alpha=\alpha_o}^2 \right\} \\ &= -E \left\{ \left(\frac{\frac{\partial f(z|\alpha)}{\partial \alpha}}{f(z|\alpha)} \right)_{\alpha=\alpha_o}^2 \right\} \\ &= -E \left\{ \left(\frac{\partial \ln f(z|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}^2 \right\}. \end{aligned} \quad (3.8)$$

Note that the derivative can be taken inside the integral since it is with respect to the parameter α and not z (this can be shown by a limit argument; see pp. 66–68 of [10]). With this information in mind, the asymptotic qualities of MLE will be derived.

To show that MLE converges in probability to the true parameter, α_o , consider the parameter values $\alpha = \alpha_o \pm \delta$ where δ is some arbitrarily small positive number. By *Jensen's inequality* [25]

$$\begin{aligned} & \int_{-\infty}^{\infty} \ln \left(\frac{f(z|\alpha_o)}{f(z|\alpha_o \pm \delta)} \right) f(z|\alpha_o) dz > 0 \\ & \int_{-\infty}^{\infty} \ln f(z|\alpha_o) f(z|\alpha_o) dz - \int_{-\infty}^{\infty} \ln f(z|\alpha_o \pm \delta) f(z|\alpha_o) dz > 0 \\ & \int_{-\infty}^{\infty} \ln f(z|\alpha_o) f(z|\alpha_o) dz > \int_{-\infty}^{\infty} \ln f(z|\alpha_o \pm \delta) f(z|\alpha_o) dz \\ & E \{ \ln f(z|\alpha_o) \} > E \{ \ln f(z|\alpha_o \pm \delta) \}. \end{aligned} \tag{3.9}$$

Invoking the strong law of large numbers³ the expectations may be written as

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \ln f(z_i|\alpha_o) > \frac{1}{n} \sum_{i=1}^n \ln f(z_i|\alpha_o \pm \delta) \\ & \ln L(\{z_i\}_1^n; \alpha_o) > \ln L(\{z_i\}_1^n; \alpha_o \pm \delta). \end{aligned} \tag{3.10}$$

This equation shows that, for almost all sample sequences $\{z_i\}_1^n$, $\ln L(\{z_i\}_1^n; \alpha_o)$ will be greater than $\ln L(\{z_i\}_1^n; \alpha_o \pm \delta)$. Since $\ln f(z|\alpha)$ is analytic by condition (ii), it is differentiable and continuous for all α so that there is a stationary point (the derivative

³If y is a random variable with finite variance, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i \xrightarrow{a.s.} E\{y\}$$

where *a.s.* means “almost surely,” or “with probability 1.”

of the log-likelihood function in Equation (3.3) is zero at this point) within the region $\alpha = \alpha_o \pm \delta$. If the stationary point is $\alpha^*(\{z_i\}_1^n) \in (\alpha_o \pm \delta)$ then by letting $\delta \rightarrow 0$, $\alpha^*(\{z_i\}_1^n) \rightarrow \alpha_o$. So MLE converges with probability 1 to a maximum of the likelihood function (3.1) which may or may not be a global maximum value.

Given the existence of a solution, one can show that this solution is asymptotically normally distributed. Using a Taylor series expansion⁴ about α_o , the true value of the parameter, Equation (3.3) may be written as (with $\alpha^*(\{z_i\}_1^n)$ simply represented by α^*)

$$\begin{aligned} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha^*} &= \\ \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o} + (\alpha^* - \alpha_o) \sum_{i=1}^n \left(\frac{\partial^2 \ln f(z_i;\alpha)}{\partial \alpha^2} \right)_{\alpha=\alpha_o} + \text{H.O.T.} \\ &= 0. \end{aligned}$$

The term H.O.T. represents higher order terms in $(\alpha^* - \alpha_o)$. Next multiply both sides of this equation by $1/n$

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o} + (\alpha^* - \alpha_o) \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 \ln f(z_i|\alpha)}{\partial \alpha^2} \right)_{\alpha=\alpha_o} + \frac{1}{n} \text{H.O.T.} = 0.$$

⁴The Taylor series expansion of $f(x)$ about the point x_o is

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(x_o)}{i!} (x - x_o)^i.$$

By setting T_1 equal to the the second summation term scaled by $1/n$ ⁵ and combining the $1/n$ scalar with the H.O.T. term to form $\overline{\text{H.O.T.}}$, this equation may be written as

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o} + (\alpha^* - \alpha_o) T_1 + \overline{\text{H.O.T.}} = 0.$$

After a few lines of algebra this equation becomes

$$k\sqrt{n}(\alpha^* - \alpha_o) = \frac{\frac{1}{k\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}}{-\left[\frac{1}{k^2} T_1 + \overline{\text{H.O.T.}} \right]} \quad (3.11)$$

(recall that k is the standard deviation of the partial derivative with respect to α of the *true* log-likelihood $\ln f(z|\alpha)$ as given in Equation (3.4)). Finally, take the limit

⁵ T_1 (“ T_1 ” is used to represent the first order term of the expansion) equals Equation (3.8),

$$E\{[(\partial^2/\partial\alpha^2)\ln f(z|\alpha)]_{\alpha=\alpha_o}\} = -E\left\{\left(\frac{\partial \ln f(z|\alpha)}{\partial \alpha}\right)_{\alpha=\alpha_o}^2\right\} \quad (3.8)$$

in the limit as n approaches infinity by the strong law of large numbers.

as the number of samples n grows to infinity

$$\begin{aligned}
\lim_{n \rightarrow \infty} k\sqrt{n}(\alpha^* - \alpha_o) &= \frac{\lim_{n \rightarrow \infty} \frac{1}{k\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}}{- \left[\frac{1}{k^2} \lim_{n \rightarrow \infty} T_1 + \lim_{n \rightarrow \infty} \overline{\text{H.O.T.}} \right]} \\
\lim_{n \rightarrow \infty} \alpha^* &= \frac{\lim_{n \rightarrow \infty} \frac{1}{k\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}}{- \lim_{n \rightarrow \infty} k\sqrt{n} \left[\frac{1}{k^2} \lim_{n \rightarrow \infty} T_1 + \lim_{n \rightarrow \infty} \overline{\text{H.O.T.}} \right]} + \lim_{n \rightarrow \infty} \alpha_o \\
\lim_{n \rightarrow \infty} \alpha^* &= \frac{\lim_{n \rightarrow \infty} \frac{1}{k^2 n} \sum_{i=1}^n \left(\frac{\partial \ln f(z_i;\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}}{- \left[\frac{1}{k^2} \lim_{n \rightarrow \infty} T_1 + \lim_{n \rightarrow \infty} \overline{\text{H.O.T.}} \right]} + \alpha_o \\
\lim_{n \rightarrow \infty} \alpha^*(\{z_i\}_i^n) &\xrightarrow{a.s.} \mathcal{N} \left\{ 0, \left(\frac{k\sqrt{n}}{k^2 n} \right)^2 \right\} + \alpha_o = \mathcal{N} \left\{ \alpha_o, \left(\frac{1}{k\sqrt{n}} \right)^2 \right\}. \quad (3.12)
\end{aligned}$$

The final steps in the derivation of Equation (3.12) requires an explanation. The term $\lim_{n \rightarrow \infty} \alpha_o$ in the second line is simply α_o since the true value of the parameter is *not* a function of the number of samples n . The denominator in the third line contains the term $\lim_{n \rightarrow \infty} T_1$ which converges to $-k^2$ by the strong law of large numbers, condition (iv), and identity (c). The other term in the denominator, $\lim_{n \rightarrow \infty} \overline{\text{H.O.T.}}$, is zero since these higher order terms are the higher order moments of z as $n \rightarrow \infty$, which are finite by condition (v), and each of these terms is scaled by $(\alpha^* - \alpha_o)^i$, $i \geq 2$, which converges to zero as $n \rightarrow \infty$. Therefore, the denominator converges to $-(-k^2/k^2 + 0) = 1$ with probability 1. Finally, the central limit theorem states that a sum of n i.i.d. random variables y_i , with finite mean $E\{y\} = \mu$ and variance $E\{(y - \mu)^2\} = \sigma^2$, converges as the number of samples becomes infinite to a normally distributed random variable with mean and variance parameters $n\mu$ and $n\sigma^2$, respectively. Applying this theorem to the third line of the equation leads to the conclusion that α^* is a normally distributed random variable with a mean of zero (by identity (b)) and a variance of $(k\sqrt{n}/(k^2 n))^2$. The last line then follows from the third line,

and the MLE estimate $\alpha^*(\{z_i\}_1^n)$ is in fact asymptotically normally distributed with probability 1.

Finally, the third asymptotic quality of MLE, efficiency, can be shown in two ways. The easiest way to show that the MLE estimate $\alpha^*(\{z_i\}_1^n)$ is asymptotically efficient is to note that its variance is inversely proportional to the number of samples n , so as this number increases, the variance decreases. In fact, the variance is zero in the asymptotic limit as $n \rightarrow \infty$. The second method to show the asymptotic efficiency of the MLE estimate is to use Cramér's definition of an asymptotically efficient estimator [10]:

$$\lim_{n \rightarrow \infty} \text{eff}(\alpha^*(\{z_i\}_1^n)) = \frac{1}{(1/k)^2 E \left\{ \left(\frac{\partial \ln f(z|\alpha)}{\partial \alpha} \right)_{\alpha=\alpha_o}^2 \right\}} = \frac{1}{k^2/k^2} = 1. \quad (3.13)$$

The efficiency function $\text{eff}(\cdot)$ is used to determine the efficiency of an estimate. By taking the limit as n approaches infinity of the efficiency function, one may calculate the asymptotic efficiency of an estimate. A value of 1 indicates that the estimate is the most efficient estimate possible. Therefore, $\alpha^*(\{z_i\}_1^n)$ is the most efficient estimate for α .

3.1.2 MLE Measure Function. The method of maximum likelihood can be related to the problem of approximating one pdf with another, which is the goal of this thesis. Again consider the problem of estimating a single pdf parameter α given a set of n i.i.d. samples. The log-likelihood function is given in Equation (3.2) and the likelihood equation is given by Equation (3.3). Since the original (true) pdf, $f(z|\alpha_o)$, is known, we need to relate this pdf to an approximate pdf based on the likelihood equation. To do so, begin with the log-likelihood equation and set the derivative equal to some number, c , when some approximated value of the parameter, $\hat{\alpha}$, is input into the equation. Next, multiply both sides of this equation by $1/n$ and take the limit as

$n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \alpha} \ln f(z_i | \alpha) \Big|_{\alpha=\hat{\alpha}} = \lim_{n \rightarrow \infty} \frac{1}{n} c = 0$$

$$\xrightarrow{a.s.} E \left\{ \left(\frac{\partial \ln f(z | \alpha)}{\partial \alpha} \right)_{\alpha=\hat{\alpha}} \right\} = 0.$$

This equation may be rewritten in integral form as

$$\int_{-\infty}^{\infty} \frac{1}{f(z | \hat{\alpha})} \left(\frac{\partial}{\partial \alpha} f(z | \alpha) \right)_{\alpha=\hat{\alpha}} f(z | \alpha_o) dx = 0 \quad (3.14)$$

where $f(z | \alpha_o)$ is the true density and $f(z | \hat{\alpha})$ is the approximate pdf. This MLE measure function may be used to evaluate the “fit” of $f(z | \hat{\alpha})$ to $f(z | \alpha_o)$.

3.2 Expectation Maximization

Generally, the solution to the likelihood equation given by Equation (3.3) may require solving nonlinear differential equations for the MLE of the parameter. In particular, for the problem of finding the parameters of a mixture density, the single likelihood equation usually becomes a set of nonlinear differential equations without an analytic solution [26]. Instead, an approximate solution is sought by an iterative approach such as Newton’s method or some form of this method, such as Rao “Scoring” [22, 26]. As an alternative to the traditional Newton-like approaches to solving the nonlinear differential equations, the Expectation Maximization (EM) algorithm is another iterative approach to solving for the MLE parameters. The EM algorithm offers a number of desirable qualities, including reliable convergence to the MLE of the parameters and computational tractability [26]. However, convergence to the solution may be slow even when applied to relatively simple problems [26].

The EM algorithm is useful when applied to MLE problems in which maximizing a “complete data” log-likelihood function is easier than maximizing the observed “incomplete data” log-likelihood function [26]. An example will be given in Subsection

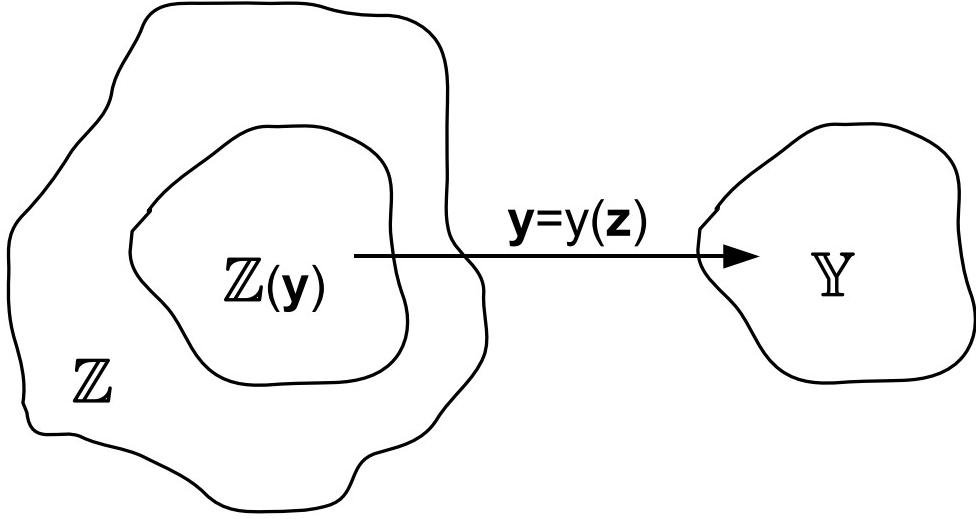


Figure 3.1: Sample spaces for the EM algorithm derivation.

[3.2.2](#) to illustrate this point. The observations may be viewed as *incomplete data* in that they are drawn from a sample space that is mapped from a subset of the *complete data* sample space [11, 26]. To clarify this point, Figure 3.1 depicts the sample space of the complete data, \mathbf{z} , as \mathbb{Z} and the sample space of the observed incomplete data, \mathbf{y} , as \mathbb{Y} . The observed incomplete data sample space is actually the mapped image of a subset of the complete data sample space, $\mathbb{Z}(\mathbf{y})$, which is mapped according to the many-to-one transformation (i.e., non-invertible) $y(\mathbf{z})$ [11].

[3.2.1](#) *Theoretical Derivation of the EM Algorithm.* The EM algorithm for observations from an exponential family (e.g., Gaussian, Poisson, and Multinomial) is derived from a theoretical perspective in this section. This derivation, including elaboration on certain details, follows that of Dempster et al. [11], who first drew widespread attention to the algorithm and developed a generalized version. The final result, Equation (3.17), is a direct consequence of the MLE approach under certain conditions and has been noted in different articles prior to the publication of [11].

The pdf of the complete data sample space is $f(\mathbf{z}|\boldsymbol{\alpha})$ and the pdf of the observed incomplete data sample space is $g(\mathbf{y}|\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is the vector set of parameters of

the pdfs. These densities are related by⁶

$$g(\mathbf{y}|\boldsymbol{\alpha}) = \int_{z \text{ not in } \mathbb{Z}(\mathbf{y})} f(\mathbf{z}|\boldsymbol{\alpha}) d\mathbf{z}. \quad (3.15)$$

Considering only exponential families of pdfs for possible candidates of $f(\cdot)$, the complete data pdf has the form

$$f(\mathbf{z}|\boldsymbol{\alpha}) = \frac{b(\mathbf{z})e^{\boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z})}}{a(\boldsymbol{\alpha})} \quad (3.16)$$

where $\mathbf{t}(\mathbf{z})$ represents the vector of *complete data sufficient statistics*⁷, $b(\mathbf{z})$ is a non-negative scalar function of \mathbf{z} , and $a(\boldsymbol{\alpha})$ is a normalization scalar [9]. The aim of the EM algorithm is to maximize $g(\mathbf{y}|\boldsymbol{\alpha})$ by appropriate choice of $\boldsymbol{\alpha}$ and by using the complete data pdf $f(\mathbf{z}|\boldsymbol{\alpha})$.

The derivation of the EM algorithm for exponential families of distributions begins by noting that the pdf of the complete data conditioned on the observed incomplete data may be written using conditional probability as [11, 26]

$$f(\mathbf{z}|\boldsymbol{\alpha}) = f(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha})f(\mathbf{y}|\boldsymbol{\alpha})$$

since \mathbf{y} is related to \mathbf{z} through the transformation $\mathbf{y} = y(\mathbf{z})$. To emphasize the distinction between the pdfs given above, let $f(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha}) = k(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha})$ and $f(\mathbf{y}|\boldsymbol{\alpha}) = g(\mathbf{y}|\boldsymbol{\alpha})$.

⁶In the paper by Dempster et al., this integral is over the domain $\mathbb{Z}(\mathbf{y})$. However, based on the written description in [11] and the corresponding graphical depiction of the sample spaces in Figure 3.1, the domain “ z not in $\mathbb{Z}(\mathbf{y})$ ” appears to be more appropriate for the purpose of this section.

⁷Heuristically, a set of observations from some underlying parent distribution is considered a sufficient statistic if it contains enough information to make a correct inference about certain properties of the parent distribution. See Chapter 2 of [42] for a more informative definition.

Now, the equation becomes⁸

$$f(\mathbf{z}|\boldsymbol{\alpha}) = k(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha})g(\mathbf{y}|\boldsymbol{\alpha}),$$

where the first term is [11]

$$k(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha}) = \frac{b(\mathbf{z})e^{\boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z})}}{a(\boldsymbol{\alpha}|\mathbf{y})}.$$

Taking the natural logarithm of this equation and rearranging terms yields an equation for the log-likelihood of $g(\mathbf{y}|\boldsymbol{\alpha})$:

$$\begin{aligned}\ln g(\mathbf{y}|\boldsymbol{\alpha}) &= \ln f(\mathbf{z}|\boldsymbol{\alpha}) - \ln k(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha}) \\ &= \ln b(\mathbf{z}) + \boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z}) - \ln a(\boldsymbol{\alpha}) - [\ln b(\mathbf{z}) + \boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z}) - \ln a(\boldsymbol{\alpha}|\mathbf{y})] \\ &= \ln a(\boldsymbol{\alpha}|\mathbf{y}) - \ln a(\boldsymbol{\alpha}).\end{aligned}$$

The $a(\cdot)$ terms can be expressed using the law of total probability:

$$\begin{aligned}\int_{\mathbb{Z}(\mathbf{y})} k(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha}) d\mathbf{z} &= 1 \\ \int_{\mathbb{Z}(\mathbf{y})} b(\mathbf{z}) e^{\boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z})} d\mathbf{z} &= a(\boldsymbol{\alpha}|\mathbf{y})\end{aligned}$$

⁸An attentive reader may question why $a(\boldsymbol{\alpha}|\mathbf{y})$ is conditioned on \mathbf{y} but neither $b(\mathbf{z})$ nor $\mathbf{t}(\mathbf{z})$ have this conditioning. The answer to this question lies in Figure 3.1. By conditioning on \mathbf{y} , \mathbf{z} is known to exist only over the restricted sample space $\mathbb{Z}(\mathbf{y})$, which will be seen in the subsequent paragraphs. So, although the conditioning is not made explicit for these terms, the conditioning has not been neglected. This notation is consistent with that of [11], so the conditioning is not explicitly represented in the aforementioned terms of the above equation.

since the conditioning on \mathbf{y} implies $\mathbf{z} \in \mathbb{Z}(\mathbf{y})$, and

$$\begin{aligned}\int_{\mathbb{Z}} f(\mathbf{z}|\boldsymbol{\alpha}) d\mathbf{z} &= 1 \\ \int_{\mathbb{Z}} b(\mathbf{z}) e^{\boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z})} d\mathbf{z} &= a(\boldsymbol{\alpha}).\end{aligned}$$

In an analogous manner to maximizing the log-likelihood function given in Equation (3.2), to maximize $\ln g(\mathbf{y}|\boldsymbol{\alpha})$, take the partial derivative with respect to the parameter vector $\boldsymbol{\alpha}$ and equate to a row vector of zeros, making the substitutions given in the last two results.

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\alpha}} \ln g(\mathbf{y}|\boldsymbol{\alpha}) &= \frac{\partial}{\partial \boldsymbol{\alpha}} \ln a(\boldsymbol{\alpha}|\mathbf{y}) - \frac{\partial}{\partial \boldsymbol{\alpha}} \ln a(\boldsymbol{\alpha}) \\ &= \frac{1}{a(\boldsymbol{\alpha}|\mathbf{y})} \frac{\partial a(\boldsymbol{\alpha}|\mathbf{y})}{\partial \boldsymbol{\alpha}} - \frac{1}{a(\boldsymbol{\alpha})} \frac{\partial a(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\ &= \frac{1}{a(\boldsymbol{\alpha}|\mathbf{y})} \int_{\mathbb{Z}(\mathbf{y})} \frac{\partial}{\partial \boldsymbol{\alpha}} b(\mathbf{z}) e^{\boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z})} d\mathbf{z} - \frac{1}{a(\boldsymbol{\alpha})} \int_{\mathbb{Z}} \frac{\partial}{\partial \boldsymbol{\alpha}} b(\mathbf{z}) e^{\boldsymbol{\alpha}^T \mathbf{t}(\mathbf{z})} d\mathbf{z} \\ &= \int_{\mathbb{Z}(\mathbf{y})} \mathbf{t}^T(\mathbf{z}) k(\mathbf{z}|\mathbf{y}, \boldsymbol{\alpha}) d\mathbf{z} - \int_{\mathbb{Z}} \mathbf{t}^T(\mathbf{z}) f(\mathbf{z}|\boldsymbol{\alpha}) d\mathbf{z} \\ &= E_{\mathbf{z}}\{\mathbf{t}^T(\mathbf{z})|\mathbf{y}, \boldsymbol{\alpha}\} - E_{\mathbf{z}}\{\mathbf{t}^T(\mathbf{z})|\boldsymbol{\alpha}\} = \mathbf{0}^T \quad (3.17)\end{aligned}$$

As with the likelihood equation given in Equation (3.3), the left-hand side of Equation (3.17) equals zero (in this case a row vector of zeros) when $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{ml}$, which implies that $E_{\mathbf{z}}\{\mathbf{t}(\mathbf{z})|\mathbf{y}, \boldsymbol{\alpha}_{ml}\} = E_{\mathbf{z}}\{\mathbf{t}(\mathbf{z})|\boldsymbol{\alpha}_{ml}\}$.

The EM algorithm follows from Equation (3.17) when it is interpreted as a two-step process. First, the complete data sufficient statistics $\mathbf{t}(\mathbf{z})$ are estimated by $\hat{\mathbf{t}}^{(i)} = E_{\mathbf{z}}\{\mathbf{t}(\mathbf{z})|\mathbf{y}, \boldsymbol{\alpha}^{(i)}\}$ at iteration i of the algorithm. Since the estimate is the expected value of the sufficient statistics conditioned on the incomplete data observation

vector \mathbf{y} , this step is referred to as the “E-step” or “expectation step.” Next, find the value of $\boldsymbol{\alpha}$ that maximizes the likelihood by invoking the equality $E_{\mathbf{z}}\{\mathbf{t}^{(i)}|\boldsymbol{\alpha}\} = \hat{\mathbf{t}}^{(i)}$ to obtain the next iteration of the estimated parameter vector $\boldsymbol{\alpha}^{(i+1)}$ (of course, an initial value must be set for $\boldsymbol{\alpha}^{(0)}$ for the algorithm to begin). This second step in the EM algorithm is referred to as the “M-step” or “maximization step.”

3.2.2 EM Algorithm Example. Now that the theory of the EM algorithm has been presented, the algorithm will be applied to a simple discrete random variable problem found in [35] which is a modified version of an example given in [11]. Consider an array of five sensors with outputs $\{z_1, z_2, z_3, z_4, z_5\}$ following a Multinomial distribution with pmf $f(\mathbf{z}|\boldsymbol{\alpha})$ ⁹; this data set and pdf correspond to the complete data. Assume that the sensor array has a malfunction such that the first and second sensor outputs are summed. Then the observed incomplete data set is $\{y_1, y_2, y_3, y_4\}$ ($y_1 = z_1 + z_2$, $y_2 = z_3$, $y_3 = z_4$, and $y_4 = z_5$) with a Multinomial pmf $g(\mathbf{y}|\boldsymbol{\alpha})$. Recall from basic probability that the sum of the realizations of the random variables of a Multinomial distribution are constrained by the total number of “trials,” n . In the case of this example, the sum of realizations from the complete data distribution must equal n_z (i.e., $n_z = z_1 + z_2 + z_3 + z_4 + z_5$). Likewise, the sum of realizations from the incomplete data distribution must equal n_y (i.e., $n_y = y_1 + y_2 + y_3 + y_4$).

This scenario is well suited for the EM algorithm since maximizing the complete data pmf, $f(\mathbf{z}|\boldsymbol{\alpha})$, is easier than maximizing the incomplete data pmf, $g(\mathbf{y}|\boldsymbol{\alpha})$. To see this point, solve for the MLE using both likelihood functions. If

$$f(\mathbf{z}|\boldsymbol{\alpha}) = \frac{(z_1 + z_2 + z_3 + z_4 + z_5)!}{z_1!z_2!z_3!z_4!z_5!} \left(\frac{1}{2}\right)^{z_1} \left(\frac{\alpha}{4}\right)^{z_2} \left(\frac{1-\alpha}{4}\right)^{z_3} \left(\frac{1-\alpha}{4}\right)^{z_4} \left(\frac{\alpha}{4}\right)^{z_5}$$

$$g(\mathbf{y}|\boldsymbol{\alpha}) = \frac{(y_1 + y_2 + y_3 + y_4)!}{y_1!y_2!y_3!y_4!} \left(\frac{1+\alpha}{2}\right)^{y_1} \left(\frac{1-\alpha}{4}\right)^{y_2} \left(\frac{1-\alpha}{4}\right)^{y_3} \left(\frac{\alpha}{4}\right)^{y_4}$$

⁹To this point, the notation $p(\cdot)$ has been used to represent a pmf. For this subsection, $f(\cdot)$ and $g(\cdot)$ are used to represent pmfs to stay consistent with the notation used in Section 3.2.

then the solutions to the likelihood equation given in Equation (3.3) for each log-likelihood function (or, equivalently, but with more algebra, each likelihood function) are¹⁰

$$(z_2 + z_3 + z_4 + z_5)\hat{\alpha}_{ml} - (z_2 + z_5) = 0$$

$$(y_1 + y_2 + y_3 + y_4)\hat{\alpha}_{ml}^2 - (y_1 - 2y_2 - 2y_3 - y_4)\hat{\alpha}_{ml} - 2y_4 = 0.$$

It is evident that solving a linear equation for $\hat{\alpha}_{ml}$ using the complete data likelihood function is simpler than solving a quadratic equation using the incomplete data likelihood function. Therefore, the EM algorithm is ideally suited for this example since it simplifies the MLE problem by introducing the complete data pmf. In fact, the solutions to these equations are

$$\begin{aligned}\hat{\alpha}_{ml} &= \frac{z_2 + z_5}{z_2 + z_3 + z_4 + z_5} \\ \hat{\alpha}_{ml} &= \frac{(y_1 - 2y_2 - 2y_3 - y_4) \pm \sqrt{(y_1 - 2y_2 - 2y_3 - y_4)^2 + 8y_4(y_1 + y_2 + y_3 + y_4)}}{2(y_1 + y_2 + y_3 + y_4)}.\end{aligned}\tag{3.18}$$

Now assume that the observation vector $\mathbf{y} = [y_1, y_2, y_3, y_4]^T = [125, 18, 20, 34]^T$ is available for computing the MLE of α . First commence with the E-step to estimate the complete data sufficient statistics $\mathbf{t}(\mathbf{z}) = \mathbf{z} = [z_1, z_2, z_3, z_4, z_5]^T$ using $\hat{\mathbf{t}}^{(i)} = \hat{\mathbf{z}}^{(i)} = E_{\mathbf{z}}\{\mathbf{z}|\mathbf{y}, \alpha^{(i)}\}$. Since $z_3 = y_2 = 18$, $z_4 = y_3 = 20$, and $z_5 = y_4 = 34$ are given by \mathbf{y} , only z_1 and z_2 need to be estimated (recall that $z_1 + z_2 = y_1 = 125$). Referring to [9], the conditional pmfs are (with the observed values substituted for

¹⁰These expressions are easily obtained by simply taking the partial derivative with respect to the parameter α for each log-likelihood function and setting the results equal to zero.

the sufficient statistics where applicable)¹¹

$$\begin{aligned}
k(z_1, z_2 | z_3, z_4, z_5, \alpha^{(i)}) &= \frac{125!}{z_1! z_2!} \left(\frac{1/2}{1/2 + \alpha^{(i)}/4} \right)^{z_1} \left(\frac{\alpha^{(i)}/4}{1/2 + \alpha^{(i)}/4} \right)^{z_2} \\
k(z_1 | z_2, z_3, z_4, z_5, \alpha^{(i)}) &= \frac{125!}{z_1!(125-z_1)!} \left(\frac{1/2}{1/2 + \alpha^{(i)}/4} \right)^{z_1} \left(\frac{\alpha^{(i)}/4}{1/2 + \alpha^{(i)}/4} \right)^{(125-z_1)} \\
k(z_2 | z_1, z_3, z_4, z_5, \alpha^{(i)}) &= \frac{125!}{z_2!(125-z_2)!} \left(\frac{\alpha^{(i)}/4}{1/2 + \alpha^{(i)}/4} \right)^{z_2} \left(\frac{1/2}{1/2 + \alpha^{(i)}/4} \right)^{(125-z_2)}.
\end{aligned}$$

Note that the quantity 125! appears in the numerators above and not 197! since the conditioning on z_3, z_4, z_5 limits the number of trials of z_1 and z_2 to $197 - 72 = 125$. These conditional pmfs are Binomial with a mean of np , where n is the number of trials and p is the probability of each event [20], so the conditional expectations are

$$\begin{aligned}
\hat{z}_1^{(i)} &= 125 \cdot \frac{1/2}{1/2 + \alpha^{(i)}/4} \\
\hat{z}_2^{(i)} &= 125 \cdot \frac{\alpha^{(i)}/4}{1/2 + \alpha^{(i)}/4}.
\end{aligned} \tag{3.19}$$

Thus the estimated sufficient statistics at step i are $\hat{\mathbf{z}}^{(i)} = [\hat{z}_1^{(i)}, \hat{z}_2^{(i)}, 18, 20, 34]^T$ and the E-step is finished.

Before proceeding with the M-step, which calculates the next iteration of the parameter estimate $\alpha^{(i+1)}$, first observe that finding the value of α that maximizes $E_{\mathbf{z}}\{\mathbf{t}(\mathbf{z})|\alpha\}$ in Equation (3.17) is equivalent to finding a stationary point of $\frac{\partial}{\partial \alpha} \ln f(\mathbf{z}|\alpha)$.

¹¹In general,

$$\begin{aligned}
f(x_1, x_2 | x_3, x_4, x_5) &= \frac{f(x_1, x_2, x_3, x_4, x_5)}{f(x_3, x_4, x_5)} \\
f(x_3, x_4, x_5) &= \frac{n!}{x_3! x_4! x_5! (n - x_3 - x_4 - x_5)!} p_3^{x_3} p_4^{x_4} p_5^{x_5} (1 - p_3 - p_4 - p_5)^{n - x_3 - x_4 - x_5} \\
f(x_1, x_2 | x_3, x_4, x_5) &= \frac{(n - x_3 - x_4 - x_5)!}{x_1! x_2!} \left(\frac{p_1}{1 - p_3 - p_4 - p_5} \right)^{x_1} \left(\frac{p_2}{1 - p_3 - p_4 - p_5} \right)^{x_2}
\end{aligned}$$

where $f(\cdot)$ is a Multinomial pmf, n is the number of trials, and p_i is the probability of z_i for $i = 1, 2, 3, 4, 5$.

This fact follows from

$$\begin{aligned} E_{\mathbf{z}}\{t(\mathbf{z})|\alpha\} &= \frac{\partial}{\partial\alpha} \ln f(\mathbf{z}|\alpha) = 0 \\ &= \frac{\partial}{\partial\alpha} (\ln b(\mathbf{z}) + \alpha t(\mathbf{z}) - \ln a(\alpha)) = 0 \\ &= t(\mathbf{z}) \end{aligned}$$

which is the equation used in the M-step. Therefore, the M-step is carried out by applying MLE using the complete data log-likelihood function $\ln f(\mathbf{z}|\alpha)$ which is

$$\alpha^{(i+1)} = \frac{\hat{z}_2^{(i)} + 34}{\hat{z}_2^{(i)} + 18 + 20 + 34} \quad (3.20)$$

(this is the first of the two MLE equations given in Equation (3.18)).

Although the previous example is relatively simple, it illustrates the main virtue of the EM algorithm. Notice that the MLE using the incomplete data log-likelihood function $\ln g(\mathbf{y}|\alpha)$ results in solving a quadratic equation (given in the second line of Equation (3.18)) while the corresponding equation for the complete data log-likelihood function is linear (the first line of Equation (3.18)). The algorithm proceeds by first computing the E-step, Equation (3.19), followed by maximization of the likelihood function in the M-step, Equation (3.20). These steps continue until the difference between the most recent successive parameter iterative estimates reaches a desired value.

3.2.3 General Form of the EM Algorithm. Although Equation (3.17) was introduced as the EM algorithm, this equation is only valid for exponential families of pdfs. A more general form of the algorithm is given in [11] which is introduced in this subsection as a bridge to the notation used in the next section. This form also enables combining the E- and M-steps. The general form of the EM algorithm defines the E-step as

$$Q(\boldsymbol{\alpha}', \boldsymbol{\alpha}^{(i)}) = E \left\{ \ln f(\mathbf{z}|\boldsymbol{\alpha}') | \mathbf{y}, \boldsymbol{\alpha}^{(i)} \right\}$$

and the M-step as

$$\boldsymbol{\alpha}^{(i+1)} = \arg \max_{\boldsymbol{\alpha}'} Q(\boldsymbol{\alpha}', \boldsymbol{\alpha}^{(i)}).$$

The form of these steps makes it possible to combine the E- and M-steps conveniently into a single step.

As an illustration, the general form of the EM algorithm may be applied to the example in the previous section with the E- and M-steps combined:

$$\begin{aligned} \frac{\partial}{\partial \alpha'} Q(\alpha', \alpha^{(i)}) &= \frac{\partial}{\partial \alpha'} \left[E \left\{ \ln \frac{(z_1 + z_2 + 18 + 20 + 34)!}{z_1! z_2! 18! 20! 34!} \middle| \alpha^{(i)} \right\} + E\{z_1 | \alpha^{(i)}\} \ln(1/2) \right. \\ &\quad \left. + E\{z_2 | \alpha^{(i)}\} \ln\left(\frac{\alpha'}{4}\right) + 38 \ln\left(\frac{1 - \alpha'}{4}\right) + 34 \ln\left(\frac{\alpha'}{4}\right) \right] \\ &= 0 + 0 + \frac{\hat{z}_2^{(i)}}{\alpha'} - \frac{38}{1 - \alpha'} + \frac{34}{\alpha'} \\ &= 0. \end{aligned}$$

Solving this equation for α' leads to the next iteration of the estimated parameter $\alpha^{(i+1)}$,

$$\alpha^{(i+1)} = \frac{\hat{z}_2^{(i)} + 34}{\hat{z}_2^{(i)} + 18 + 20 + 34},$$

which is the same as Equation (3.20).

3.3 Multivariate Gaussian Mixture Estimation

This section applies the theory behind MLE and the EM algorithm to the problem of estimating a multivariate Gaussian mixture pdf. This problem is motivated by the goal of the thesis, which is to represent one Gaussian mixture with another containing a reduced number of components. Although this section pertains to estimating a multivariate Gaussian mixture pdf, the techniques involved may provide valuable insight into approximating one Gaussian mixture with another. The asymptotic representation of the estimates are of particular interest since they relate the true, full-component mixture to the approximated, lower-component mixture.

In [26], Redner and Homer derive the MLE and EM algorithm equations for estimating the mean, covariance, and mixture weights of a multivariate Gaussian mixture pdf. Their derivation includes the constraint that the mixture weights are non-negative and sum to one, but the authors acknowledge that the constraints on the covariance being symmetric and positive definite are not explicitly imposed. However, the forms of the solutions are claimed to uphold the qualities of the initial quantities: if the initial mixture weights are non-negative and sum to one then so do the estimated mixture weights. Likewise, if the initial covariance is positive definite and symmetric, then the estimated covariance has the same properties (see pp. 217–218 of [26]).

When reading the following sections, one should keep in mind the context of the derivations of the parameter estimation equations. First, the parameters to be estimated are from the multivariate Gaussian mixture pdf given in Equation (2.23). The resulting likelihood function, provided N i.i.d. sample *vector* observations, is

$$\begin{aligned} L(\{\mathbf{z}_i\}_1^N; \boldsymbol{\Omega}) &= \prod_{i=1}^N f(\mathbf{z}_i | \boldsymbol{\Omega}) \\ &= \prod_{i=1}^N \sum_{j=1}^M p_i f(\mathbf{z}_i | \boldsymbol{\mu}_j, \mathbf{P}_j). \end{aligned} \quad (3.21)$$

Again, it is emphasized that $\{\mathbf{z}_i\}_1^N = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$ is a *vector* set of i.i.d. observations in which each *vector* in the set has dimension m (see Equation (2.24)); that is, $\mathbf{z}_i = [z_{k1}, \dots, z_{km}]^T$ for $i = 1, \dots, N$. Also note that the constraints on the parameters are the same as those mentioned in Section 2.3. Second, the MLE approach will maximize this equation only using the constraint that the mixture weights are non-negative and sum to one. Third, for the EM algorithm method to be used, Redner and Homer suppose that the sample observations are incomplete in that they are not “labeled” (it is unknown which mixture component spawned which samples) and propose a complete data sample $\mathbf{y}_i = (\mathbf{z}_i, \mathbf{l}_i)$ where \mathbf{l}_k is the vector containing the labels for each z_{uv} (i.e., for each component sample of each vector set of observations).

3.3.1 MLE of a Multivariate Gaussian Mixture. The MLE of the parameters of a multivariate Gaussian mixture log-likelihood function,

$$\begin{aligned}\ln L(\{\mathbf{z}_i\}_1^N; \boldsymbol{\Omega}) &= \sum_{i=1}^N \ln f(\mathbf{z}_i | \boldsymbol{\Omega}) \\ &= \sum_{i=1}^N \ln \sum_{j=1}^M p_j f(\mathbf{z}_i | \boldsymbol{\mu}_j, \mathbf{P}_j),\end{aligned}$$

are the solutions to the likelihood equations,

$$\nabla_{\boldsymbol{\Omega}} \ln L(\{\mathbf{z}_i\}_1^N; \boldsymbol{\Omega}) \Big|_{\boldsymbol{\Omega}=\hat{\boldsymbol{\Omega}}_{ml}} = \mathbf{0},$$

in the unconstrained optimization case ¹². After applying the constraint on the mixture weights, the MLE of this parameter is given by (with the ml subscript suppressed) [26]¹³

$$\hat{p}_j = \frac{1}{N} \sum_{i=1}^N \hat{p}_j \frac{f(\mathbf{z}_i | \boldsymbol{\mu}_j, \mathbf{P}_j)}{f(\mathbf{z}_i | \boldsymbol{\Omega})} \quad , \quad (3.22)$$

for $j = 1, \dots, M$. Next, the MLE of the mixture component means and covariances are found as the solutions to the unconstrained likelihood equations

$$\nabla_{\boldsymbol{\mu}_j} \ln L(\{\mathbf{z}_i\}_1^N; \boldsymbol{\Omega}) \Big|_{\boldsymbol{\Omega}=\hat{\boldsymbol{\Omega}}_{ml}} = \mathbf{0}$$

$$\nabla_{\mathbf{P}_j} \ln L(\{\mathbf{z}_i\}_1^N; \boldsymbol{\Omega}) \Big|_{\boldsymbol{\Omega}=\hat{\boldsymbol{\Omega}}_{ml}} = \mathbf{0}$$

¹²In general, the vector derivative of a scalar function dependent on the components of the derivative is given by [35]

$$\nabla_{\mathbf{x}} F(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} F(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_N} F(\mathbf{x}) \end{bmatrix}.$$

¹³One troubling aspect of this equation is that it is in terms of itself! That is, the mixture weight appears on both sides of the equation. This apparent flaw is not noted in [26]. Later, the MLE equations will be set aside in favor of the EM algorithm equations, so this apparent mathematical contradiction is avoided.

for $j = 1, \dots, M$. These solutions are¹⁴

$$\begin{aligned}\hat{\boldsymbol{\mu}}_j &= \frac{\sum_{i=1}^N \mathbf{z}_i \hat{p}_j \frac{f(\mathbf{z}_i | \hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)}{f(\mathbf{z}_i | \hat{\boldsymbol{\Omega}})}}{\sum_{i=1}^N \hat{p}_j \frac{f(\mathbf{z}_i | \hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)}{f(\mathbf{z}_i | \hat{\boldsymbol{\Omega}})}} \\ \hat{\mathbf{P}}_j &= \frac{\sum_{i=1}^N (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_j)(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_j)^T \hat{p}_j \frac{f(\mathbf{z}_i | \hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)}{f(\mathbf{z}_i | \hat{\boldsymbol{\Omega}})}}{\sum_{i=1}^N \hat{p}_j \frac{f(\mathbf{z}_i | \hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)}{f(\mathbf{z}_i | \hat{\boldsymbol{\Omega}})}}\end{aligned}\tag{3.23}$$

for $j = 1, \dots, M$.

The solutions for the MLE of the mixture weights, means, and covariances of a multivariate Gaussian mixture have some nice properties. Equation (3.22) shows that each \hat{p}_j will be non-negative since all of the quantities are non-negative, including the pdfs, and that the set sums to one. Furthermore, it can be seen that the covariance parameter solution, Equation (3.23), is a sum of symmetric, rank one matrices scaled by positive numbers which will produce a symmetric, positive *semi-definite* matrix [34]. However, Redner and Homer claim that the covariance solution is positive definite as long as the initial covariance estimate is positive definite [26]. The authors' claim may be substantiated in practice since the possibility of a sample vector \mathbf{z}_i equalling $\hat{\boldsymbol{\mu}}_j$ (thus making the covariance parameter solution positive *semi-definite*) is unlikely. These properties meet the constraints given above so that the parameters estimated by Equations (3.22) and (3.23) will result in a valid multivariate Gaussian mixture solution.

¹⁴To derive these results apply the dell operator to the unconstrained likelihood equations, use the chain and product rules, use the vector and matrix derivative identities [8, 35]

- (a) $\nabla_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = 2\mathbf{A} \mathbf{x}$
 - (b) $\nabla_{\mathbf{A}} \mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} = -\mathbf{A}^{-T} \mathbf{x} \mathbf{x}^T \mathbf{A}^{-T}$
 - (c) $\nabla_{\mathbf{A}} \det \mathbf{A} = (\det \mathbf{A}) \mathbf{A}^{-T}$
- and solve for the parameters.

Although the above estimates appear to meet the requisite constraints on the parameters, they possess at least one troubling property - each solution is in terms of itself. In Equation (3.22) the \hat{p}_j terms cancel so that an estimate cannot be made. Also, the mean and covariance solutions in Equation (3.23) are complicated non-linear relations of the parameters to be estimated.

One means of circumventing this problem (which was introduced in Section 3.2) is to implement the EM algorithm. By doing so, the estimation equations remain the same but the estimates are now iterative and thus in terms of the future and current estimates. This simplification is a direct result of applying the EM algorithm to an MLE problem where maximizing the likelihood function that is produced by the algorithm is easier than maximizing the MLE likelihood function.

3.3.2 EM Algorithm for a Multivariate Gaussian Mixture. As mentioned above, the EM algorithm can simplify the MLE solution for the parameters of a multivariate Gaussian mixture if the estimation problem can be posed in terms of complete and incomplete data (see Section 3.2). In this instance, the simplification is not the form of the parameters estimate solutions in Equations (3.22) and (3.23), but the iterative nature of the EM algorithm.

The EM algorithm is derived by modifying the MLE problem according to the developments in Section 3.2 and using the general form of the EM algorithm given in Subsection 3.2.3 rather than the result given for exponential families of distributions. To begin, specify the complete data set $\mathbf{y}_i = (\mathbf{z}_i, \mathbf{l}_i)$ where \mathbf{l}_i is the vector containing the labels for each z_{uv} as mentioned in the beginning of this section. The complete,

incomplete, and conditional likelihood functions are then

$$\begin{aligned} f(\{\mathbf{y}_i\}_1^N | \boldsymbol{\Omega}) &= \prod_{i=1}^N p_{ij} f(\mathbf{z}_i | \boldsymbol{\mu}_{ij}, \mathbf{P}_{ij}) \\ g(\{\mathbf{z}_i\}_1^N | \boldsymbol{\Omega}) &= \prod_{i=1}^N f(\mathbf{z}_i | \boldsymbol{\Omega}) \\ k(\{\mathbf{y}_i\}_1^N | \{\mathbf{z}_i\}_1^N, \boldsymbol{\Omega}) &= \frac{f(\{\mathbf{y}_i\}_1^N | \boldsymbol{\Omega})}{g(\{\mathbf{z}_i\}_1^N | \boldsymbol{\Omega})}. \end{aligned}$$

Applying the E-step yields [26] (s is the iteration index)

$$\begin{aligned} Q(\boldsymbol{\Omega}', \boldsymbol{\Omega}^{(s)}) &= E\{\ln f(\{\mathbf{y}_i\}_1^N | \boldsymbol{\Omega}') | \{\mathbf{z}_i\}_1^N, \boldsymbol{\Omega}^{(s)}\} \\ &= \sum_{j=1}^M \left[\sum_{i=1}^N \frac{p_j^{(s)} f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})} \right] \ln p_j' + \sum_{j=1}^M \sum_{i=1}^N \ln f(\mathbf{z}_i | \boldsymbol{\mu}_j', \mathbf{P}_j') \frac{p_j^{(s)} f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})} \end{aligned}$$

and applying the M-step¹⁵ produces the EM algorithm solutions for the parameters:

$$\begin{aligned} p_j^{(s+1)} &= \frac{1}{N} \sum_{i=1}^N p_j^{(s)} \frac{f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})} \\ \boldsymbol{\mu}_j^{(s+1)} &= \frac{\sum_{i=1}^N \mathbf{z}_i p_j^{(s)} \frac{f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})}}{\sum_{i=1}^N p_j^{(s)} \frac{f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})}} \\ \mathbf{P}_j^{(s+1)} &= \frac{\sum_{i=1}^N (\mathbf{z}_i - \boldsymbol{\mu}_j^{(s+1)}) (\mathbf{z}_i - \boldsymbol{\mu}_j^{(s+1)})^T p_j^{(s)} \frac{f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})}}{\sum_{i=1}^N p_j^{(s)} \frac{f(\mathbf{z}_i | \boldsymbol{\mu}_j^{(s)}, \mathbf{P}_j^{(s)})}{f(\mathbf{z}_i | \boldsymbol{\Omega}^{(s)})}} \end{aligned} \tag{3.24}$$

¹⁵To obtain this result, use the constrained maximization for the mixture weights and the unconstrained maximization for the means and covariances as in Subsection 3.3.1. The derivatives should be with respect to the primed variables.

for $j = 1, \dots, M$. Note that all of the properties mentioned in Subsection 3.3.1 apply to these solutions, but since the estimates are iterative, they are not in terms of themselves as with the MLE solutions.

3.3.3 Asymptotic Representation of the EM Algorithm. The EM algorithm estimates for the parameters of a multivariate Gaussian mixture can be asymptotically extended to include the true mixture pdf. This action is motivated by finding a method of comparing a full-component multivariate Gaussian mixture with one having a reduced number of components. To see this point, take the limit as the number of sample vectors approaches infinity and invoke the strong law of large numbers. In the equations that follow, Ω_o is the full-component or true value of the parameter set, N is the number of sample vectors, and *a.s.* means *almost surely* as previously defined.

$$\begin{aligned} \lim_{N \rightarrow \infty} p_j^{(s+1)} &= p_j^{\star,(s+1)} \xrightarrow{\text{a.s.}} \int_{\mathbf{z} \in \mathbb{Z}} p_j^{\star,(s)} \frac{f(\mathbf{z} | \boldsymbol{\mu}_j^{\star,(s)}, \mathbf{P}_j^{\star,(s)})}{f(\mathbf{z} | \boldsymbol{\Omega}^{\star,(s)})} f(\mathbf{z} | \boldsymbol{\Omega}_o) d\mathbf{z} \\ \lim_{N \rightarrow \infty} \boldsymbol{\mu}_j^{(s+1)} &= \boldsymbol{\mu}_j^{\star,(s+1)} \xrightarrow{\text{a.s.}} \frac{\int_{\mathbf{z} \in \mathbb{Z}} \mathbf{z} p_j^{\star,(s)} \frac{f(\mathbf{z} | \boldsymbol{\mu}_j^{\star,(s)}, \mathbf{P}_j^{\star,(s)})}{f(\mathbf{z} | \boldsymbol{\Omega}^{\star,(s)})} f(\mathbf{z} | \boldsymbol{\Omega}_o) d\mathbf{z}}{\int_{\mathbf{z} \in \mathbb{Z}} p_j^{\star,(s)} \frac{f(\mathbf{z} | \boldsymbol{\mu}_j^{\star,(s)}, \mathbf{P}_j^{\star,(s)})}{f(\mathbf{z} | \boldsymbol{\Omega}^{\star,(s)})} f(\mathbf{z} | \boldsymbol{\Omega}_o) d\mathbf{z}} \quad (3.25) \\ \lim_{N \rightarrow \infty} \mathbf{P}_j^{(s+1)} &= \mathbf{P}_j^{\star,(s+1)} \\ &\xrightarrow{\text{a.s.}} \frac{\int_{\mathbf{z} \in \mathbb{Z}} (\mathbf{z} - \boldsymbol{\mu}_j^{\star,(s+1)}) (\mathbf{z} - \boldsymbol{\mu}_j^{\star,(s+1)})^T p_j^{\star,(s)} \frac{f(\mathbf{z} | \boldsymbol{\mu}_j^{\star,(s)}, \mathbf{P}_j^{\star,(s)})}{f(\mathbf{z} | \boldsymbol{\Omega}^{\star,(s)})} f(\mathbf{z} | \boldsymbol{\Omega}_o) d\mathbf{z}}{\int_{\mathbf{z} \in \mathbb{Z}} p_j^{\star,(s)} \frac{f(\mathbf{z} | \boldsymbol{\mu}_j^{\star,(s)}, \mathbf{P}_j^{\star,(s)})}{f(\mathbf{z} | \boldsymbol{\Omega}^{\star,(s)})} f(\mathbf{z} | \boldsymbol{\Omega}_o) d\mathbf{z}} \end{aligned}$$

for $j = 1, \dots, M$.

Of particular interest is the relationship between the asymptotic EM Equations (3.25) and the MLE measure function given by Equation (3.14). If the MLE measure function is written in terms of a multivariate Gaussian mixture, then it has the form

$$\int_{\mathbf{z} \in \mathbb{Z}} \frac{1}{f(\mathbf{z}|\hat{\boldsymbol{\Omega}})} (\nabla_{\boldsymbol{\Omega}} f(\mathbf{z}|\boldsymbol{\Omega}))_{\boldsymbol{\Omega}=\hat{\boldsymbol{\Omega}}} f(\mathbf{z}|\boldsymbol{\Omega}_o) d\mathbf{z} = \mathbf{0}. \quad (3.26)$$

By setting $\nabla_{\boldsymbol{\Omega}} = \{\nabla_{\boldsymbol{\mu}_j}, \nabla_{\mathbf{P}_j}\}$ and solving for the mean and covariance parameters, respectively, one obtains the corresponding equations in Equation (3.25).

3.4 Summary

This chapter explored sample observation-based pdf estimation using MLE and an iterative implementation of MLE called the EM algorithm to gain insight into possible methods for approximating a multivariate Gaussian mixture pdf with one containing a lower number of mixture components. Three asymptotic qualities of MLE (convergence to the true parameter value, convergence to a Gaussian distribution, and efficiency) were derived to emphasize the effectiveness of MLE as a pdf estimation technique. The asymptotic nature of MLE is important because the MLE measure function derived in Equation (3.26) could be used to discriminate between a full-component target state Gaussian mixture pdf and an approximate reduced-component mixture pdf. The EM algorithm makes use of *complete data* to simplify the MLE problem as illustrated by the example provided in Subsection 3.2.2 in which a quadratic solution equation was replaced by a linear solution equation. This algorithm was applied to the problem of estimating the parameters of a multivariate Gaussian mixture pdf when provided with sample vector observations in Subsection 3.3.2. It may be possible to use this approach to produce an approximation to a multivariate Gaussian mixture pdf by generating samples from the original pdf and implementing Equation (3.24) to estimate the parameters of the reduced-component mixture pdf.

IV. Approximating Gaussian Mixtures & Mixture Reduction Algorithms

In the context of the tracking with measurement origin uncertainty problem outlined in Chapter II, the Bayesian solution to the problem results in a Gaussian mixture representation of the target state pdf. When new measurements are received at each scan, the number of mixture components, which represent hypotheses about each measurement in the entire measurement history with regard to the overall target state, increases. The rate of increase is usually exponential and the tracking problem quickly becomes computationally intractable.

Approximating the full-component target state Gaussian mixture pdf at the end of every measurement processing cycle is one means of remedying this problem. There are generally two types of mixture approximations that are used in practice. The first method of approximation is to reduce the mixture to a single component, such as in the PDA and JPDA algorithms [30, 31, 38]. However, this method is a rather crude approximation to the original mixture, and it ignores well-spaced mixture components, potentially losing valuable information at the end of the scan cycle [30]. The second approach approximates the full-component mixture pdf with one containing a lower number of components, as in Williams' recently developed Integral Square Error (ISE) cost-function-based mixture reduction algorithm [38, 40, 41] and others [27, 30, 31, 32].

The focus of this chapter is approximating a full-component target state Gaussian mixture pdf with one having a lower number of components based on some mathematical measure. This type of approximation reduces the number of mixture components by either merging or deleting existing ones based on the measure of each action. Section 4.1 introduces the various measures used to indicate the goodness of fit of a low-order approximate Gaussian mixture to the original target state Gaussian mixture pdf. Next, Section 4.2 presents two heuristic algorithms, the Greedy algorithms, for choosing the “best” mixture reduction actions from a pool of proposed reductions based on the output of a measure function. Sections 4.3 and 4.4 cover two existing mixture reduction algorithms (MRAs), Salmond’s Joining and Clustering al-

gorithms and Williams' ISE cost-function-based algorithm, respectively, for optimally reducing the number of mixture components.

4.1 Measure Functions for Gaussian Mixture Approximation

In the context of this thesis, a measure function used for Gaussian mixture approximation may be categorized as either a *true* distance measure or a *pseudo*-distance measure. The distinction between the two classes of distance measures is purely mathematical, since the second type of measure function does not satisfy the *triangle inequality*¹, but maintains the non-negative and, in some cases, the symmetry properties of a true distance. A measure function is applied to Gaussian mixture pdf approximation as a means of discriminating between two pdfs, such as a Gaussian mixture pdf and a reduced-component approximation of the same. In this capacity, a measure function provides a criterion for mixture reduction decisions and is a key element of an MRA.

All of the distance measure functions presented in this section exhibit at least one of the three properties of a true distance measure:

1. Symmetry
2. Satisfying the *triangle inequality*
3. Non-negativeness.

Symmetry, as it pertains to distance measures, means that the distance between two pdfs is independent of the order in which the distance is calculated. For instance, suppose that there are two vectors a and b which originate from a common point

¹ The *triangle inequality* states that the length of the difference vector between two vectors a and b is less than or equal to the sum of the lengths of the two vectors [10]. That is,

$$\|a - b\| \leq \|a\| + \|b\|.$$

Another form of the *triangle inequality* is [34]

$$\|a + b\| \leq \|a\| + \|b\|.$$

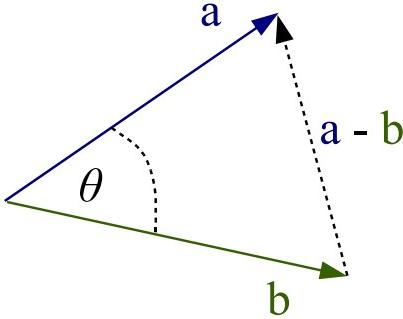


Figure 4.1: Two vectors having a common origin in the space in which they exist.

(such as the origin of the space in which they exist) as shown in Figure 4.1. One possible symmetric distance measure between the two vectors is the norm of the difference vector, since $\|a - b\| = \|b - a\|$. If a measure function satisfies the *triangle inequality*, then it is a candidate as a true distance measure [17]. To write a more profound statement about this property would be unwise, given the limited mathematical background of this author. However, the *triangle inequality* imposes a nice physical constraint of a distance since the “real-world” concept of distance also has this property². Non-negativeness imposes another attractive physical constraint on a true distance measure because it ensures that the measured distance is *never* negative (although it may be zero).

Throughout this section, Gaussian mixture pdfs may be thought of as infinite-dimensional vectors in *Hilbert space*. The length of a pdf in *Hilbert space* is mathematically defined as the square root of the inner product of the pdf with itself (the norm as defined in *Hilbert space*) [37]. If the original multivariate Gaussian mixture pdf given its weight, mean, and covariance parameters is represented by $f(\mathbf{x}|\boldsymbol{\Omega}_o)$,

²A basic “real-world” example of the *triangle inequality* is to measure the length of two sticks, and then “add” the two sticks (using the second form of the *triangle inequality*) by placing them end to end and measure their combined length. Common sense dictates that the combined length of the sticks not exceed the separate length of each stick.

then its length is

$$\begin{aligned}\|f(\mathbf{x}|\boldsymbol{\Omega}_o)\| &\triangleq \sqrt{\langle f(\mathbf{x}|\boldsymbol{\Omega}_o), f(\mathbf{x}|\boldsymbol{\Omega}_o) \rangle} \\ &= \sqrt{\int_{-\infty}^{\infty} f^2(\mathbf{x}|\boldsymbol{\Omega}_o) d\mathbf{x}}\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the inner product notation. Similarly, the length of a reduced-component approximation of a multivariate Gaussian mixture pdf given its weight, mean, and covariance parameters is represented by

$$\begin{aligned}\|f(\mathbf{x}|\hat{\boldsymbol{\Omega}})\| &\triangleq \sqrt{\langle f(\mathbf{x}|\hat{\boldsymbol{\Omega}}), f(\mathbf{x}|\hat{\boldsymbol{\Omega}}) \rangle} \\ &= \sqrt{\int_{-\infty}^{\infty} f^2(\mathbf{x}|\hat{\boldsymbol{\Omega}}) d\mathbf{x}}.\end{aligned}$$

In both of these equations, $\boldsymbol{\Omega}$ represents the weight, mean, and covariance parameters of the mixtures, and the subscript o or “hat” notation indicates that the parameters are for the original mixture or the approximate mixture, respectively.

4.1.1 True Distance Measures. A true distance measure satisfies the three properties of a true distance introduced in Section 4.1. The following true distance measures are presented in this subsection:

- Kolmogorov Variational Distance (L_1 Distance, Total Variation Distance)
- Integral Square Error cost function
- Hellinger Distance
- Correlation Measure
- Hellinger Affinity Measure (Bhattacharyya coefficient).

The first three true distance measures may be thought of geometrically as measuring *length of the error vector* between some function of two pdfs, while the last two true

distance measures may be viewed as measuring the *cosine of the angle* between some function of two pdfs (as in Figure 4.1).

4.1.1.1 Kolmogorov Variational Distance. The Kolmogorov Variational Distance (or L_1 Distance) is given by (with the notation T indicating that this measure is a true distance measure) [1, 2, 24, 28, 38]

$$T_K\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} = \int_{-\infty}^{\infty} |f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega})| d\mathbf{x}. \quad (4.1)$$

The Kolmogorov Variational Distance may be viewed as the sum of the absolute differences between the uncountably infinite, infinitesimally small elements of the two pdfs. This measure clearly adheres to properties 1 and 3 of a true distance measure because

$$T_K\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} = T_K\{f(\mathbf{x}|\hat{\Omega}), f(\mathbf{x}|\Omega_o)\}$$

and the absolute value function ensures that the distance is non-negative. To show that this measure also meets the *triangle inequality*, let $T_K\{f(\mathbf{x}|\Omega_o), 0\}$ represent $\|a\|$ and $T_K\{f(\mathbf{x}|\hat{\Omega}), 0\}$ represent $\|b\|^3$. Then, $T_K\{f(\mathbf{x}|\Omega_o) - 0, f(\mathbf{x}|\hat{\Omega}) - 0\}$ represents $\|a - b\|$, and the *triangle inequality* is satisfied since

$$\begin{aligned} T_K\{f(\mathbf{x}|\Omega_o) - 0, f(\mathbf{x}|\hat{\Omega}) - 0\} &= \int_{-\infty}^{\infty} |f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega})| d\mathbf{x} \\ &\leq \int_{-\infty}^{\infty} |f(\mathbf{x}|\Omega_o) - 0| d\mathbf{x} + \int_{-\infty}^{\infty} |f(\mathbf{x}|\hat{\Omega}) - 0| d\mathbf{x} \\ &= T_K\{f(\mathbf{x}|\Omega_o), 0\} + T_K\{f(\mathbf{x}|\hat{\Omega}), 0\}. \end{aligned}$$

Equation (4.1) seems difficult to evaluate without approximation. For instance, the absolute value breaks the integral into two (not necessarily contiguous) pieces:

³Note that the distance between a non-zero vector and the zero vector is simply the length of the non-zero vector *using the given distance measure*.

one in which $f(\mathbf{x}|\Omega_o) > f(\mathbf{x}|\hat{\Omega})$ and the other in which $f(\mathbf{x}|\Omega_o) < f(\mathbf{x}|\hat{\Omega})$. Since the pdfs are Gaussian, the integral over a portion of the domain of the pdf would be extremely difficult to evaluate without some sort of approximation [38] such as a numerical approximation of an integral. Assuming that such an approximation is computationally expensive relative to the update time of the tracking system, this measure function appears unsuitable for practical real-time implementation.

4.1.1.2 Integral Square Error Cost Function. Williams' Integral Square Error (ISE) cost function demonstrated the best performance against a single target in heavy clutter tracking when implemented in an MRA for a Bayesian tracking algorithm in the presence of measurement origin uncertainty [38]. His cost function is

$$\begin{aligned} T_{ISE}\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \langle f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}), f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}) \rangle \\ &= \int_{-\infty}^{\infty} \left[f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}) \right]^2 d\mathbf{x} \\ &= \int_{-\infty}^{\infty} \left[f^2(\mathbf{x}|\Omega_o) + f^2(\mathbf{x}|\hat{\Omega}) - 2f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) \right] d\mathbf{x} \end{aligned} \quad (4.2)$$

where the first term in the last line of this equation is the original mixture *self-likeness* term, the second term is the reduced mixture *self-likeness* term, and the third term is the *cross-likeness* term [38, 40, 41]. The ISE cost function is a true distance measure since it meets the requisite properties listed in Section 4.1. Properties 1 and 2 hold because interchanging the pdfs does not change the measure (it is symmetrical) and the squaring function insures the distance measure is non-negative. It satisfies the *triangle inequality* since (using similar representations for $\|a\|$, $\|b\|$, and $\|a - b\|$ as

before)

$$\begin{aligned}
T_{ISE}\{f(\mathbf{x}|\Omega_o) - 0, f(\mathbf{x}|\hat{\Omega}) - 0\} &= \int_{-\infty}^{\infty} \left[f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}) \right]^2 d\mathbf{x} \\
&= \int_{-\infty}^{\infty} \left[f^2(\mathbf{x}|\Omega_o) + f^2(\mathbf{x}|\hat{\Omega}) \right] d\mathbf{x} - \int_{-\infty}^{\infty} 2f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) d\mathbf{x} \\
&\leq \int_{-\infty}^{\infty} \left[f^2(\mathbf{x}|\Omega_o) + f^2(\mathbf{x}|\hat{\Omega}) \right] d\mathbf{x} \\
&= \int_{-\infty}^{\infty} [f(\mathbf{x}|\Omega_o) - 0]^2 d\mathbf{x} + \int_{-\infty}^{\infty} \left[f(\mathbf{x}|\hat{\Omega}) - 0 \right]^2 d\mathbf{x} \\
&= T_{ISE}\{f(\mathbf{x}|\Omega_o), 0\} + T_{ISE}\{f(\mathbf{x}|\hat{\Omega}), 0\}.
\end{aligned}$$

Unlike the Kolmogorov Variational Distance, an exact closed form solution exists for the ISE cost function when the pdfs are multivariate Gaussian mixtures, so this true distance measure is well-suited for real-time application as the reduction decision criterion in an MRA [38, 40, 41]. By observing that the first line in Equation (4.2) is the squared length of the point-by-point difference between two pdfs, the ISE cost function may be interpreted using the *Hilbert space* vector analogy as the squared length of the error vector in Figure 4.1. From this perspective, it should not be surprising that the ISE cost function has been shown to provide the best single target in heavy clutter tracking performance to date when implemented as the MRA for a Bayesian tracking in clutter algorithm [38, 40, 41]. For the same area of discrepancy between the two densities, $f(\mathbf{x}|\Omega_o)$ and $f(\mathbf{x}|\hat{\Omega})$, a high narrow area of discrepancy is weighted more severely by this cost measure than a low broad discrepancy region, unlike the case of using the Kolmogorov Variational Distance [38].

It is interesting to point out that the *cross-likeness* term by itself is used as a measure in [15]. In this paper, the authors suggest that the Expected Likelihood Kernel could be used for discrimination between two pdfs. This measure is (where M is used to distinguish this measure from a true distance measure)

$$M_E\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} = \int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})d\mathbf{x} \quad (4.3)$$

which is one-half of the *cross-likeness* term in the ISE cost function, Equation (4.2). Although this measure has an exact closed-form solution, it is not immediately clear how to use this measure to make mixture reduction decisions⁴. For instance, if it is used as a measure of the distance between a pdf and an approximation of this pdf, one would like to select the approximation that produces the smallest value of Equation (4.3) (i.e., the smallest distance). However, the concept of orthogonality in mathematics would indicate that if Equation (4.3) evaluates to zero for two pdfs, then the pdfs are very dissimilar. So, small values for this measure would mean that the two pdfs are not similar. For this reason, the Expected Likelihood Kernel is not categorized as a distance measure, nor is it clear how to apply this measure to mixture reduction⁵.

4.1.1.3 Hellinger Distance. The Hellinger Distance is similar to the ISE cost function, except it operates on the square root of the pdfs rather than directly on the pdfs. Several variations of the Hellinger Distance are found in the

⁴A more detailed discussion of the suitability of the Expected Likelihood Kernel as an appropriate measure function is provided in Section 5.1.

⁵Although the Expected Likelihood Kernel appears as a component in the ISE cost function (it is the *cross-likeness* term) and the Correlation Measure (presented in Subsection 4.1.1.4), it is not conceptually related to either of these true distance measure functions, and outputs of the Expected Likelihood Kernel do not have the physical interpretation of a distance.

literature [15, 24, 25], and in this thesis, the Hellinger Distance is defined as [24]

$$\begin{aligned}
T_H\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \sqrt{\frac{1}{2}\left\langle \sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})}, \sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})} \right\rangle} \\
&= \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} \left[\sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})} \right]^2 d\mathbf{x}} \\
&= \sqrt{1 - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}}. \tag{4.4}
\end{aligned}$$

Given the similarity between the Hellinger Distance and ISE cost function, it is readily apparent that the Hellinger Distance meets the first and third requirements of a true distance. As expected, it also satisfies the *triangle inequality*:

$$\begin{aligned}
T_H\{f(\mathbf{x}|\Omega_o) - 0, f(\mathbf{x}|\hat{\Omega}) - 0\} &= \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} \left[\sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})} \right]^2 d\mathbf{x}} \\
&= \sqrt{1 - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}} \\
&\leq 1 \\
&= \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o) d\mathbf{x} + \frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\hat{\Omega}) d\mathbf{x}} \\
&= T_H\{f(\mathbf{x}|\Omega_o), 0\} + T_H\{f(\mathbf{x}|\hat{\Omega}), 0\}.
\end{aligned}$$

Equation (4.4) shows that the Hellinger Distance has the potential to be a promising alternative to the ISE cost function since it does not require the computation of the two *self-likeness* terms in Equation (4.2). However, the integral of the square root of a product of multivariate Gaussian mixture pdfs is *extremely* difficult to evaluate exactly in closed form [15, 38]. Placing numerical evaluation aside, the

Hellinger Distance could be *approximated* in closed form using a truncated binomial series or the “heuristic approximation” of replacing the square root of a sum with a sum of square roots proposed in [15].

4.1.1.4 Correlation Measure. The Correlation Measure is the first of two true distance measures which may conceptually be thought of as calculating the *cosine of the angle* between the original and approximate reduced-component Gaussian mixture pdfs as if they were vectors in *Hilbert space* (see Figure 4.1)⁶. It is written as [19]

$$\begin{aligned} T_C\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \frac{\langle f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega}) \rangle}{\sqrt{\langle f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\Omega_o) \rangle \langle f(\mathbf{x}|\hat{\Omega}), f(\mathbf{x}|\hat{\Omega}) \rangle}} \\ &= \frac{\int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o) f(\mathbf{x}|\hat{\Omega}) d\mathbf{x}}{\sqrt{\int_{-\infty}^{\infty} f^2(\mathbf{x}|\Omega_o) d\mathbf{x} \int_{-\infty}^{\infty} f^2(\mathbf{x}|\hat{\Omega}) d\mathbf{x}}}. \end{aligned} \quad (4.5)$$

This measure is used in the field of communication systems as a means of determining the similarity between two signals [19]. Again, using a vector analogy, one can easily see that if two signals (vectors) are the same, then the angle between the two signals (vectors) is zero and the cosine of this angle is one. If the two signals (vectors) are perpendicular, then they are considered “indifferent” to each other or independent, and the cosine of the angle between the two signals (vectors) is zero [19]. Finally, if the two signals (vectors) point in opposite directions, then the angle between them is 180°, and the signals (vectors) are completely different [19]. Returning to Equation (4.5) and noting that the two pdfs *never* assume negative quantities,

⁶The cosine of the angle between two vectors a and b in *any* dimensional Euclidean space is given by (using the more general inner product notation) [34]

$$\cos \theta = \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle \langle b, b \rangle}}.$$

the range of possible values of the right-hand side of this equation is restricted to be between zero and one. Thus, if one seeks a good approximation of a Gaussian mixture pdf, then the Correlation Measure between the original and approximation mixtures would have to be close to one.

Showing that Equation (4.5) is a true distance measure requires a deviation from the previous pattern of proofs. The Correlation Measure is symmetric because interchanging the arguments of this distance measure does not affect the distance calculation, and it is non-negative since it is restricted to values between zero and one, as noted in the previous paragraph. In addition, this measure *must* satisfy the *triangle inequality* since the numerator of Equation (4.5) must be less than or equal to the denominator by the *Schwartz inequality* (with equality), and the *triangle inequality* reduces to the *Schwartz inequality* [34]. Thus the Correlation Measure has the quality of being a true distance.

4.1.1.5 Hellinger Affinity Measure. The Hellinger Affinity Measure [24] (or Bhattacharyya coefficient) is the second true distance measure which may be viewed as the *cosine of the angle* between the square root of the original and the square root of the approximate reduced-component Gaussian mixture pdfs using the *Hilbert space* vector analogy⁷. This measure was considered for application to Gaussian mixture reduction by Lainiotis and Park in [18], and it has the form

$$\begin{aligned} T_A\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \langle \sqrt{f(\mathbf{x}|\Omega_o)}, \sqrt{f(\mathbf{x}|\hat{\Omega})} \rangle \\ &= \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}. \end{aligned} \quad (4.6)$$

Notice that Equation (4.6) is simply the cosine of the angle between the element-wise square root of each vector representing the two pdfs in *Hilbert space*. This insight

⁷In this case, the square root of the infinite-dimensional vectors representing each pdf is taken as the element-wise square root since the square root of a vector is undefined.

is apparent when one considers replacing the pdfs in the Correlation Measure with the square root of each pdf. Then, the denominator would reduce to one because the integral of a pdf over the entire sample space of the random variable it describes is one. The Hellinger Affinity Measure satisfies the three requirements of a true distance, which can be shown in a similar manner as that for the Correlation Measure, and it suffers from the same difficulty in finding an exact closed-form solution as the Hellinger Distance (their functional forms are clearly related, as seen by comparing Equations (4.6) and (4.4)).

4.1.2 Pseudo-Distance Measures. Pseudo-distance measures are different from true distance measures because they do not satisfy all three properties of a true distance listed in Section 4.1. However, this distinction should not preclude their application to pdf discrimination problems. In fact, the two Kullback-Leibler measures are probably used more in practice than all of the true distance measures combined [12, 15, 24]. The following pseudo-distance measures are presented in this subsection:

- Kullback-Leibler Mean Information
- Kullback-Leibler Divergence
- Salmond’s Joining Algorithm cost function.

The Kullback-Leibler measures were originally derived for use in sample observation-based problems in which the true pdf of the parent distribution (i.e., the distribution that spawned the samples) is unknown, but the measures may also be applied when the true pdf is known exactly (as is the case for this thesis).

4.1.2.1 Kullback-Leibler Mean Information. When applied to a sample observation-based problem, one interpretation of the Kullback-Leibler Mean Information measure is that it provides an indication of the amount of new information gained for discrimination between two pdfs by making an additional observation [17]. As a pseudo-distance measure, the Kullback-Leibler Mean Information measure is

not symmetric and it does not satisfy the *triangle inequality* [17]. However, by using *Jensen's inequality* (see Equation (3.9)) one can show that this distance measure is non-negative [24]. The Kullback-Leibler Mean Information is (the D standing for distance to distinguish pseudo-distance measures from true distance measures)

$$D_{MI}\{f(\mathbf{x}|\boldsymbol{\Omega}_o), f(\mathbf{x}|\hat{\boldsymbol{\Omega}})\} = \int_{-\infty}^{\infty} f(\mathbf{x}|\boldsymbol{\Omega}_o) \ln \frac{f(\mathbf{x}|\boldsymbol{\Omega}_o)}{f(\mathbf{x}|\hat{\boldsymbol{\Omega}})} d\mathbf{x}. \quad (4.7)$$

The natural logarithm in Equation (4.7) makes obtaining an exact closed-form solution of this distance *extremely* difficult when the pdfs are Gaussian mixtures [38].

4.1.2.2 Kullback-Leibler Divergence. The Kullback-Leibler Divergence measure provides a sense of the difficulty in discriminating between two pdfs [17]. Like the Mean Information, the Divergence is non-negative, but unlike this measure, the Divergence is symmetric [17, 24]. The Kullback-Leibler Divergence is given by

$$D_D\{f(\mathbf{x}|\boldsymbol{\Omega}_o), f(\mathbf{x}|\hat{\boldsymbol{\Omega}})\} = \int_{-\infty}^{\infty} [f(\mathbf{x}|\boldsymbol{\Omega}_o) - f(\mathbf{x}|\hat{\boldsymbol{\Omega}})] \ln \frac{f(\mathbf{x}|\boldsymbol{\Omega}_o)}{f(\mathbf{x}|\hat{\boldsymbol{\Omega}})} d\mathbf{x}. \quad (4.8)$$

Again, the presence of the natural logarithm function makes finding an exact closed-form solution *extremely* difficult when the pdfs are Gaussian mixtures [38].

4.1.2.3 Salmond's Joining Algorithm Cost Function. Salmond based the cost function for his Joining Algorithm on penalizing changes to the structure of the original Gaussian mixture caused by reduction actions [30]. The cost function is derived by first setting the covariance of the full-component mixture equal to the covariance of the reduced-component mixture, thus maintaining the “structure” of the original mixture. In the final step of his derivation, Salmond develops a scalar cost function based on the Mahalanobis distance [30, 31]. This cost function meets the symmetric and non-negative properties of a true distance measure, and so it is classified as a pseudo-distance measure.

Salmond's Joining Algorithm⁸ cost function is (breaking from the previous notation to maintain consistency with Salmond's notation)

$$d_{ij}^2 = \frac{p_i p_j}{p_i + p_j} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \mathbf{P}^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \quad (4.9)$$

where d_{ij}^2 is the squared distance resulting from merging mixture components i and j , and \mathbf{P} is the overall covariance of the mixture [31]. The distances between components are pair-wise compared using this cost function, and the components that fall below some distance threshold are merged [31]. Salmond's Joining Algorithm will be discussed further in Section 4.3.

4.2 Greedy Algorithms for the Assignment Problem

Consider the problem of reducing a multivariate Gaussian mixture pdf with $N_H(k)$ mixture components to one with a reduced number $N_R(k)$, $N_H(k) > N_R(k)$, by either merging or deleting components. A pool of potential reduction actions for the original mixture pdf is formed by proposing the deletion of any one of the mixture components, $1, \dots, N_H(k)$, or the merging of every pair of distinct mixture components, where the merged component's parameters are calculated using Equation (2.26). After all possible actions are proposed, this pool consists of $N_H(k)$ and $N_H(k)[N_H(k)-1]/2$ ⁹ proposed mixture component deletion and merge actions, respectively. Selecting the "best" reduction action or actions from the set of $N_H(k) + \{N_H(k)[N_H(k)-1]/2\}$ possible actions may be viewed as an assignment problem, and either of two heuristic algorithms, called Greedy Algorithm A and Greedy Algorithm B¹⁰, may be applied

⁸A description of Salmond's Joining and Clustering algorithms is given in Section 4.3.

⁹The number of possible merge actions is equivalent to selecting two mixture components from $N_H(k)$ mixture components without replacement and without regard to order. Thus, the number of possible merge actions is [20]

$$\binom{N_H(k)}{2} = \frac{N_H(k)!}{2![N_H(k)-2]!} = \frac{N_H(k)[N_H(k)-1][N_H(k)-2]!}{2[N_H(k)-2]!} = \frac{N_H(k)[N_H(k)-1]}{2}.$$

¹⁰Williams' ISE cost-function-based algorithm makes use of Greedy Algorithm B [41].

to this problem [29]. Both of these algorithms assign the “best” reduction action or actions based on the output of a measure function.

Again, consider the problem of reducing an $N_H(k)$ -component Gaussian mixture pdf to an $N_R(k)$ -component mixture pdf. According to Greedy Algorithm A, one selects the $[N_H(k) - N_R(k)]$ reduction actions from $N_H(k) + \{N_H(k)[N_H(k) - 1]/2\}$ possible actions with the most favorable outputs of some measure function [29]. In the case of the distance and pseudo-distance measure functions of Section 4.1, the $[N_H(k) - N_R(k)]$ best reduction actions are those with the smallest corresponding distance measures. The best $[N_H(k) - N_R(k)]$ reduction actions are then executed, and a reduced-component Gaussian mixture pdf approximation is obtained.

In contrast, Greedy Algorithm B selects the $N_H(k) - N_R(k)$ best actions through an *iterative* process. At the first iteration, a set of $\{N_H(k) + N_H(k)[N_H(k) - 1]/2\}$ possible mixture reduction actions is proposed. Then, Greedy Algorithm B selects the best reduction action from the set based on some measure function, and executes this reduction action by either deleting the selected mixture component or merging two selected mixture components. At the next iteration, a new set of possible deletion and merge actions is proposed. Since the number of mixture components was reduced by one in the previous iteration, the number of possible mixture reduction actions is now $[N_H(k) - 1] + \{[N_H(k) - 1][N_H(k) - 2]/2\}$. This process continues until the desired number of reductions, $[N_H(k) - N_R(k)]$, is obtained.

4.3 Salmond’s Joining & Clustering Algorithms

The Joining and Clustering algorithms are two MRAs for a Bayesian tracking in clutter algorithm. In [30, 31], Salmond develops and tests these algorithms for a scenario of a single target in clutter. The algorithms are applied separately to the task of reducing the number of mixture components of the *a posteriori* target state pdf while attempting to maintain track on the target.

Before developing his algorithms, Salmond created design objectives which are paraphrased below [31].

- (i) The pdf approximations resulting from the algorithms should be the same form as the approximated pdf; i.e., the approximation should be a multivariate Gaussian mixture.
- (ii) The algorithm allows the user to specify the number of components in the approximation.
- (iii) Reduction actions will be guided by a predetermined threshold on a cost function that measures the change to the “structure” of the approximated mixture. Reduction should continue until either the threshold is breached or the number of specified components is reached. This cost criterion is chosen since it is computationally tractable¹¹.
- (iv) Intuitively, the approximation should maintain the overall mean and covariance of the original mixture.
- (v) The algorithms should be computationally efficient such that the approximation can take place before the completion of a scan cycle¹².

Pair-wise component merging, carried out by Equation (2.26), is the focus of each iteration of the Joining Algorithm [31]. This process is governed by requirement (ii) from above so that merging continues until either a preset threshold is breached or the requisite number of reduced components is met [31]. The cost function is given in Equation (4.9) and the threshold T is set by noting that the cost function is bounded below the dimension, n , of the target state random process vector, and simulation results indicate that $T = 0.001n$ [31].

¹¹This part of requirement (iii) appears to have been the motivation behind Williams’ thesis [38]. He suggested that a more robust cost function could be utilized and still be computationally feasible given the improvement in computing power since Salmond’s dissertation.

¹²The completion of one scan cycle means that the propagation and measurement update stages (as given in Subsection 2.2.1) have been completed.

The Clustering Algorithm [31] iteratively finds the component with the largest mixture weight and groups other components which are closest to this *cluster center* (the component with the largest mixture weight at each iteration). Grouping decisions are based on a cost function with the same form as (4.9) but with slightly different components [31]:

$$D_i^2 = \frac{p_i p_c}{p_i + p_c} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_c)^T \mathbf{P}_c^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_c) \quad (4.10)$$

(the subscript c indicates that the parameters are from the cluster center component). As with the Joining Algorithm, a threshold is used to determine when components should be clustered, but unlike the Joining Algorithm, this threshold is linearly increased if the preset number of reduced components is not achieved at the end of a clustering iteration [31]. The value of this threshold has a nice geometric interpretation as being the volume of the hyperellipsoid centered about the mean of the cluster center and encompassing a certain percentage of the components' probability mass [31].

4.4 Williams' ISE Cost-Function-Based Algorithm

In [38], Williams develops the ISE Initialization and ISE Iterative Optimization MRAs. The ISE Initialization MRA uses the ISE cost function to choose a starting point for the approximate reduced-component mixture parameters. This starting point is then used in the ISE Iterative Optimization MRA which optimizes the values of the mixture parameters using the gradient of the ISE cost function with respect to each of the three multivariate mixture parameters. However, it was noted in [38] that the value of the mixture parameters obtained using the Initialization algorithm provided a starting point which typically was close to the Iteratively Optimized parameter values, so that the added computational load of the optimization algorithm could be avoided by using the Initialization algorithm only [38].

The ISE Initialization MRA combines the benefits of the pdf measure functions used by Alspach, Lainiotis, and Park [1, 2, 18] and adapts the work of Salmond [31]

to create an MRA that produces a track life that surpasses that of the Joining and Clustering algorithms [40] when considering a tracking scenario of a single target in heavy clutter. Conceptually, the ISE cost function given by Equation (4.2) is similar to the Kolmogorov Variational Distance used by Alspach and the Hellinger Affinity Measure used by Lainiotis and Park in that it takes into consideration the *entire* pdf of the original and reduced-component mixtures. However, unlike the Kolmogorov Variational Distance and Hellinger Affinity Measure, the ISE cost function does not need to be approximated to obtain a closed-form solution in the case of multivariate Gaussian mixtures [38, 40, 41]. Williams adopts all of the requirements posed in [31] (listed in Section 4.3), but requirement (iii) is modified to remove thresholding as a criterion to stop reduction. The merging Equations (2.26) developed in [31] are used to combine mixture components but clustering is avoided [38].

A flowchart summarizing the ISE Initialization MRA (which uses Greedy Algorithm B) is shown in Figure 4.2 (reproduced from [38]). Like the Joining Algorithm, each pair of components is potentially merged using (2.26). The cost of these merging actions is evaluated using the ISE cost function which reduces to (keeping most of Williams' original notation)

$$\begin{aligned} J_{HR} &= \sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_R(k)} p_i \bar{p}_j \mathcal{N}\{\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_j, \mathbf{P}_i + \bar{\mathbf{P}}_j\} \\ J_{RR} &= \sum_{i=1}^{N_R(k)} \sum_{j=1}^{N_R(k)} \bar{p}_i \bar{p}_j \mathcal{N}\{\bar{\boldsymbol{\mu}}_i | \bar{\boldsymbol{\mu}}_j, \bar{\mathbf{P}}_i + \bar{\mathbf{P}}_j\} \\ J_{HH} &= \sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_H(k)} p_i p_j \mathcal{N}\{\boldsymbol{\mu}_i | \boldsymbol{\mu}_j, \mathbf{P}_i + \mathbf{P}_j\} \end{aligned}$$

where the bars over the parameters indicate those of the reduced mixture, $N_H(k)$ and $N_R(k)$ are the number of original and reduced mixture components, respectively, at the end of scan k , and $\mathcal{N}\{\cdot\}$ is a multivariate Gaussian pdf with the specified

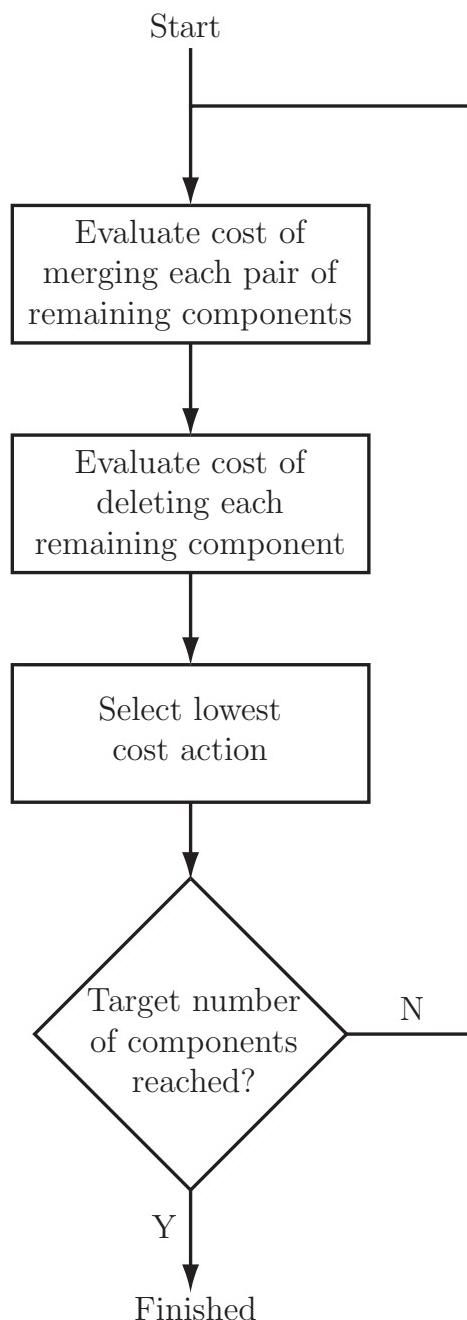


Figure 4.2: A flowchart for the ISE cost-function-based Initialization algorithm (which uses Greedy Algorithm B).

functional arguments in brackets¹³ [38, 40, 41]. Unlike the Joining Algorithm, the cost is also evaluated for deleting each component. Given all of the potential reduction actions, the one with the lowest cost is executed. This process continues until the preset number of reduced mixture components is met (as illustrated in Figure 4.2), in contrast to Salmond’s Joining Algorithm which may continue merging components below this number as long as the threshold is not exceeded [31].

4.5 Summary

In Chapter II, the Bayesian solution for tracking a target in clutter leads to a Gaussian mixture target state pdf in which the number of mixture components (hypotheses) grows without bound over time. Implementing this solution requires some type of mixture reduction algorithm (MRA) to limit the number of hypotheses to a manageable number while maintaining good tracking performance. The choice of measure function used as a criterion to make mixture reduction decisions has a major impact on tracking performance because it is a key element of an MRA. True distance and pseudo-distance measure functions were defined, and a comparison of the geometric interpretation of each distance measure was made (when applicable), along with a judgment of the measure function’s suitability for practical implementation. The task of reducing the number of components of a multivariate Gaussian mixture pdf by utilizing a measure function as the reduction decision criterion is an assignment problem. Two heuristic assignment algorithms, Greedy Algorithm A and Greedy Algorithm B, were presented as candidate solutions to this problem. Two existing MRAs, Salmond’s Joining and Clustering algorithms and Williams’ ISE cost-function-based algorithm, were summarized because they form the basis for this thesis.

¹³This notation may be a little confusing since the mean, which is a deterministic vector, is used as the random vector of the pdf. The $\mathcal{N}\{\cdot\}$ terms should be interpreted as a functional form and *not* an actual pdf.

V. Gaussian Mixture Reduction Algorithm Development & Analysis

Several mixture reduction algorithms (MRAs) are developed in this chapter¹. The MRAs are composed of two essential pieces: the measure function used as a decision criterion for selecting mixture reduction actions, and the assignment algorithm which uses the measure function to decide which actions to execute. The most suitable measure functions from Chapter III and Chapter IV are mated with one of the assignment algorithms of Chapter IV to produce each new MRA. Various univariate Gaussian mixture pdfs are used to test the new MRAs, and the results are analyzed to identify the best candidates for implementation in a full-scale Bayesian tracking algorithm for use in the presence of measurement origin uncertainty.

5.1 Measure Function & Assignment Algorithm Selection

Before developing an MRA, a suitable measure function must be chosen and coupled with either Greedy Algorithm A or Greedy Algorithm B from Section 4.2. A measure function is considered suitable if it can be exactly evaluated or approximately evaluated in closed-form when the pdfs are multivariate Gaussian mixtures, and if the interpretation of its results is unambiguous. Either Greedy assignment algorithm may be used with any suitable measure function. Salmond's Joining Algorithm cost function is not considered because it does not incorporate the reduction action effects on the entire target state pdf. However, Williams thoroughly compared his Integral Square Error cost function with Salmond's Joining Algorithm cost function in [38,40, 41].

Excluding Salmond's Joining Algorithm cost function, the nine measure functions presented in Chapters III and IV will be considered for implementation in new MRAs:

1. MLE measure (Equation (3.26))

¹The sample-based multivariate Gaussian mixture approximation method using the EM algorithm suggested in Chapter III is not considered in this chapter.

2. Kolmogorov Variational Distance (Equation (4.1))
3. Integral Square Error (ISE) cost function (Equation (4.2))
4. Expected Likelihood Kernel (Equation (4.3))
5. Hellinger Distance (Equation (4.4))
6. Correlation Measure (Equation (4.5))
7. Hellinger Affinity Measure (Equation (4.6))
8. Kullback-Leibler Mean Information (Equation (4.7))
9. Kullback-Leibler Divergence (Equation (4.8)).

A number of these measure functions are unsuitable. Exact or approximate closed-form evaluation of measure functions 2, 8, and 9 is extremely difficult, if not impossible, to obtain when the arguments of these functions are Gaussian mixture pdfs, as noted in the literature [15, 38], so these measure functions are not suitable. The MLE measure function is discarded for the same reason. As pointed out in Subsection 4.1.1.2, the interpretation of results generated by the Expected Likelihood Kernel is somewhat ambiguous, so this measure function is also deemed unsuitable.

To see this point, consider a case in which the Expected Likelihood Kernel of two pdfs is small. This result would imply that the two pdfs are dissimilar based on the concept of orthogonality. However, a large measure value does not necessarily correspond to a high-degree of similarity between the two pdfs under consideration. Figure 5.1 shows two five-component univariate Gaussian mixture pdfs which differ only by their respective mixture weights. The parameters for the first mixture pdf (the solid trace in the figure) and the second mixture pdf (the dash-dotted trace in the figure) are shown in Table 5.1. The Expected Likelihood Kernel of the solid-trace mixture pdf with itself is 0.075159, while the same measure between the solid-trace and dash-dotted trace pdfs is 0.075484, which is larger. Thus, in general, a larger Expected Likelihood Kernel value does not always imply a better match between pdfs.

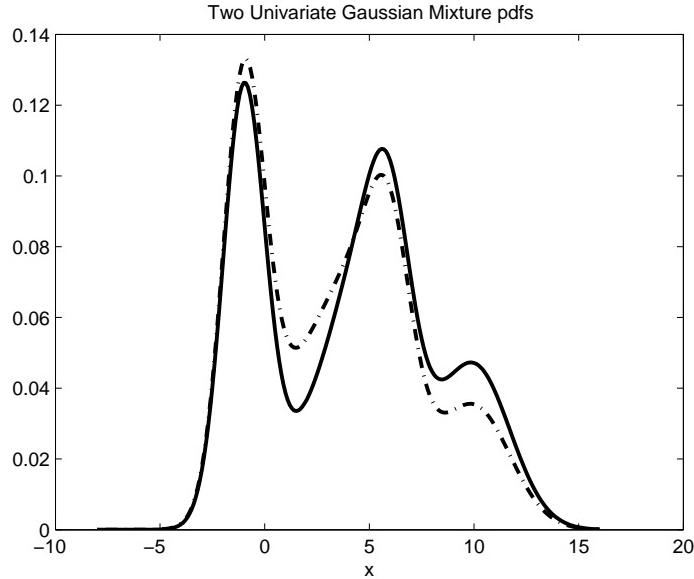


Figure 5.1: Two distinct five-component univariate Gaussian mixture pdfs with the same mean and variance parameters, but with different mixture weight parameters. The Expected Likelihood Kernel measure of the solid-trace mixture with itself is 0.075159, while the same measure between the solid-trace and dash-dotted trace mixture pdfs is 0.075484.

Parameter	First pdf	Second pdf
Weight	$[0.3, 0.1, 0.1, 0.2, 0.3]^T$	$[0.3, 0.2, 0.1, 0.15, 0.25]^T$
Mean	$[-1, 2, 6, 10, 5]^T$	$[-1, 2, 6, 10, 5]^T$
Variance	$[1, 4, 1, 3, 3]^T$	$[1, 4, 1, 3, 3]^T$

Table 5.1: Mixture parameters for the two univariate Gaussian mixture pdfs in Figure 5.1.

Measure functions 3, 5, 6, and 7 are considered suitable for implementation in an MRA. Small evaluated values from the ISE cost function and Hellinger Distance indicate that two pdfs under consideration are well-matched, while values close to one imply that two pdfs are similar when using the Correlation Measure or Hellinger Affinity Measure². An exact closed-form solution for the ISE cost function for multivariate Gaussian mixture pdfs was derived in [38, 40, 41], and an exact closed-form solution for the Correlation Measure also exists since the basic terms used in this measure function are also used in the ISE cost function. Although exact closed-form evaluations of the Hellinger Distance and Hellinger Affinity Measure may not exist for multivariate Gaussian mixture pdfs [15, 38], approximate closed-form representations of these measure functions may be found using a truncated binomial series approximation or the “heuristic approximation” cited in [15].

Of all the possible pairings of measure functions with assignment algorithms, the following set of new MRAs will be investigated:

1. ISE cost function mated with Greedy Algorithm A
2. Correlation Measure mated with Greedy Algorithm B
3. Hellinger Distance mated with Greedy Algorithm B
4. Hellinger Affinity Measure mated with Greedy Algorithm B.

The first MRA is a modification to Williams’ ISE cost-function-based MRA since it replaces Greedy Algorithm B with Greedy Algorithm A, and it is termed the ISE Shotgun MRA. The second new MRA is created by replacing the ISE cost function with the Correlation Measure and modifying the decision logic to select the measure closest to one as opposed to the one closest to zero. This new MRA is the Correlation Measure (CM) MRA. The final two new MRAs, the Hellinger Distance (HD) and Hellinger Affinity Measure (HA) MRAs, are implemented using either the truncated

²In Subsection 4.1.1, the Correlation Measure and Hellinger Affinity Measure were described as the *cosine of the angle* between two vectors in *Hilbert space*. A measure value close to one indicates that the angle between the vectors representing the pdfs (using the *Hilbert space* vector analogy) is nearly zero, so the two pdfs are similar.

binomial series approximation or the “heuristic approximation” in the next section. The performance of each new MRA will be compared to Williams’ ISE MRA, and a final comparison of the best-performing *new* MRA will be made with the ISE MRA.

5.2 *Mixture Reduction Algorithm Development*

This section presents the complete development of the four new MRAs listed at the end of the previous section. Exact closed-form solutions for the ISE cost function and Correlation Measure in the case of multivariate Gaussian mixture pdfs are shown, and approximate closed-form results for the Hellinger Distance and Hellinger Affinity Measure are derived using two different approximations. The process of proposing deletion and merge reduction actions is introduced. Finally, the selected measure functions are mated to their corresponding assignment algorithm and detailed descriptions of the four new MRAs are given.

5.2.1 Closed-Form Solutions of Select Measure Functions. Exact closed-form solutions for the ISE cost function and Correlation Measure, and approximate closed-form solutions for the Hellinger Distance and Hellinger Affinity Measure are derived in this subsection. In [38], Williams demonstrated that the product of two multivariate Gaussian pdfs results in another multivariate Gaussian pdf, and this result is used to evaluate his ISE cost function. In [15], the authors present the general probability product kernel for Gaussian pdfs which provides a generalized version of the result found in [38]. These two results are used to develop the “heuristic approximation” suggested in [15] for the Hellinger Distance and Hellinger Affinity Measure. A binomial series representation for the square root function in the two Hellinger measures is used to obtain the other approximate closed-form solution for these measures.

5.2.1.1 *ISE Cost Function & Correlation Measure Closed-Form Solutions.* Since the ISE cost function,

$$\begin{aligned}
T_{ISE}\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \langle f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}), f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}) \rangle \\
&= \int_{-\infty}^{\infty} \left[f(\mathbf{x}|\Omega_o) - f(\mathbf{x}|\hat{\Omega}) \right]^2 d\mathbf{x} \\
&= \int_{-\infty}^{\infty} \left[f^2(\mathbf{x}|\Omega_o) + f^2(\mathbf{x}|\hat{\Omega}) - 2f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) \right] d\mathbf{x} \\
&= \langle f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\Omega_o) \rangle + \langle f(\mathbf{x}|\hat{\Omega}), f(\mathbf{x}|\hat{\Omega}) \rangle - 2\langle f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega}) \rangle,
\end{aligned} \tag{4.2}$$

and the Correlation Measure,

$$\begin{aligned}
T_C\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \frac{\langle f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega}) \rangle}{\sqrt{\langle f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\Omega_o) \rangle \langle f(\mathbf{x}|\hat{\Omega}), f(\mathbf{x}|\hat{\Omega}) \rangle}} \\
&= \frac{\int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) d\mathbf{x}}{\sqrt{\int_{-\infty}^{\infty} f^2(\mathbf{x}|\Omega_o) d\mathbf{x} \int_{-\infty}^{\infty} f^2(\mathbf{x}|\hat{\Omega}) d\mathbf{x}}}, \tag{4.5}
\end{aligned}$$

share the two *self-likeness* terms and *cross-likeness* term, the required evaluations of the exact closed-form solution when the pdfs are multivariate Gaussian mixtures are the same. Although Williams' closed-form result may be applied directly to both measures, the general probability product kernel for multivariate Gaussian pdfs given in [15] is used instead for two reasons. First, the general probability product kernel is needed to derive the “heuristic approximation” for the Hellinger measures in the next subsection. Second, the result produced by the general probability product kernel may be checked against Williams' solution to validate it (at least in one case).

The general probability product kernel for two multivariate Gaussian pdfs and general ρ is given as (note that $|\cdot| = \det(\cdot)$) [15]

$$\begin{aligned}
K_\rho\{f(\mathbf{x}|\boldsymbol{\mu}_o, \mathbf{P}_o), f(\mathbf{x}|\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}})\} &= \langle f^\rho(\mathbf{x}|\boldsymbol{\mu}_o, \mathbf{P}_o), f^\rho(\mathbf{x}|\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}}) \rangle \quad (5.1) \\
&= (2\pi)^{(1-2\rho)n/2} (\rho)^{-n/2} |\mathbf{P}^\dagger|^{1/2} |\mathbf{P}_o|^{-\rho/2} |\hat{\mathbf{P}}|^{-\rho/2} \\
&\quad \cdot \exp \left\{ -\frac{\rho}{2} \left[\boldsymbol{\mu}_o^T \mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\boldsymbol{\mu}}^T \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\dagger T} \mathbf{P}^\dagger \boldsymbol{\mu}^\dagger \right] \right\} \\
\mathbf{P}^\dagger &= \left(\mathbf{P}_o^{-1} + \hat{\mathbf{P}}^{-1} \right)^{-1} \\
\boldsymbol{\mu}^\dagger &= \mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}}.
\end{aligned}$$

(Note that \mathbf{P}^\dagger is an *inverse* covariance and not a covariance.) Recall that a multivariate Gaussian mixture pdf has the form

$$f(\mathbf{x}|\Omega) = \sum_{i=1}^M p_i f(\mathbf{x}|\boldsymbol{\mu}_i, \mathbf{P}_i) \quad (2.23)$$

so that the original, full-component single-target state multivariate Gaussian mixture pdf may be written as

$$f(\mathbf{x}(k)|\Omega_o(k)) = \sum_{i=1}^{N_H(k)} p_{o,i} f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,i}, \mathbf{P}_{o,i}) \quad (5.2)$$

and the approximated, reduced-component single-target state multivariate Gaussian mixture pdf is

$$f(\mathbf{x}(k)|\hat{\Omega}(k)) = \sum_{j=1}^{N_R(k)} \hat{p}_j f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j). \quad (5.3)$$

Notice that the time dependence on k was added since the single-target state is modeled as a random process vector. At sample k , the random process vector $\mathbf{x}(k)$ is simply a multivariate Gaussian mixture vector with parameters given in $\Omega_o(k)$ or $\hat{\Omega}(k)$. In contrast, the time dependence is not explicitly shown in the mixture component parameters $p_{o,i}$, $\boldsymbol{\mu}_{o,i}$, $\mathbf{P}_{o,i}$, \hat{p}_j , $\hat{\boldsymbol{\mu}}_j$, and $\hat{\mathbf{P}}_j$ to enhance readability, but

it should be understood that these values generally change from sample to sample. Substituting Equations (5.2) and (5.3) into Equation (5.1), and setting $\rho = 1$ yields the closed-form solution for the *cross-likeness* term

$$\begin{aligned} \langle f(\mathbf{x}(k)|\boldsymbol{\Omega}_o(k)), f(\mathbf{x}(k)|\hat{\boldsymbol{\Omega}}(k)) \rangle &= \\ &\sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_R(k)} p_{o,i} \hat{p}_j K_1 \{ f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,i}, \mathbf{P}_{o,i}), f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j) \}. \end{aligned} \quad (5.4)$$

Likewise, the full-component mixture *self-likeness* term is

$$\begin{aligned} \langle f(\mathbf{x}(k)|\boldsymbol{\Omega}_o(k)), \mathbf{x}(k)|\boldsymbol{\Omega}_o(k) \rangle &= \\ &\sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_H(k)} p_{o,i} p_{o,j} K_1 \{ f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,i}, \mathbf{P}_{o,i}), f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,j}, \mathbf{P}_{o,j}) \} \end{aligned} \quad (5.5)$$

and the reduced-component mixture *self-likeness* term is

$$\begin{aligned} \langle f(\mathbf{x}(k)|\hat{\boldsymbol{\Omega}}(k)), f(\mathbf{x}(k)|\hat{\boldsymbol{\Omega}}(k)) \rangle &= \\ &\sum_{i=1}^{N_R(k)} \sum_{j=1}^{N_R(k)} \hat{p}_i \hat{p}_j K_1 \{ f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_i, \hat{\mathbf{P}}_i), f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j) \}. \end{aligned} \quad (5.6)$$

Finally, for $\rho = 1$, Equation (5.1) evaluates to³

$$\begin{aligned}
K_1\{f(\mathbf{x}|\boldsymbol{\mu}_o, \mathbf{P}_o), f(\mathbf{x}|\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}})\} &= (2\pi)^{-n/2} |\mathbf{P}_o^{-1} + \hat{\mathbf{P}}^{-1}|^{-1/2} |\mathbf{P}_o|^{-1/2} |\hat{\mathbf{P}}|^{-1/2} \\
&\cdot \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\mu}_o^T \mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\boldsymbol{\mu}}^T \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}} \right. \right. \\
&\quad \left. \left. - (\mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}})^T (\mathbf{P}_o^{-1} + \hat{\mathbf{P}}^{-1})^{-1} (\mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}}) \right] \right\} \\
&= |2\pi(\mathbf{P}_o + \hat{\mathbf{P}})|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\mu}_o^T \mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\boldsymbol{\mu}}^T \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_o^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\mathbf{P}} \mathbf{P}_o^{-1} \boldsymbol{\mu}_o \right. \right. \\
&\quad \left. \left. - \hat{\boldsymbol{\mu}}^T \hat{\mathbf{P}}^{-1} \mathbf{P}_o (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\mathbf{P}} \mathbf{P}_o^{-1} \boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}}^T \hat{\mathbf{P}}^{-1} \mathbf{P}_o (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_o^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\boldsymbol{\mu}} \right] \right\} \\
&= |2\pi(\mathbf{P}_o + \hat{\mathbf{P}})|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\mu}_o^T (\mathbf{P}_o^{-1} - \mathbf{P}_o^{-1} \hat{\mathbf{P}} (\mathbf{P}_o + \hat{\mathbf{P}})^{-1}) \boldsymbol{\mu}_o \right. \right. \\
&\quad \left. \left. + \hat{\boldsymbol{\mu}}^T (\hat{\mathbf{P}}^{-1} - \hat{\mathbf{P}}^{-1} \mathbf{P}_o (\mathbf{P}_o + \hat{\mathbf{P}})^{-1}) \hat{\boldsymbol{\mu}} - 2\boldsymbol{\mu}_o^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\boldsymbol{\mu}} \right] \right\} \\
&= |2\pi(\mathbf{P}_o + \hat{\mathbf{P}})|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\mu}_o^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \boldsymbol{\mu}_o - \boldsymbol{\mu}_o^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\boldsymbol{\mu}} \right. \right. \\
&\quad \left. \left. - \hat{\boldsymbol{\mu}}^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \boldsymbol{\mu}_o + \hat{\boldsymbol{\mu}}^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} \hat{\boldsymbol{\mu}} \right] \right\} \\
&= |2\pi(\mathbf{P}_o + \hat{\mathbf{P}})|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}})^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} (\boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}}) \right\} \tag{5.7}
\end{aligned}$$

which matches the result found in [38]. Thus, the ISE cost function and Correlation Measure are fully specified by Equations (5.4), (5.5), and (5.6), and the probability

³This result is derived by using the following linear algebra properties/relations [8, 21, 34]:

- (a) $|\mathbf{A}|^{-1} = \frac{1}{|\mathbf{A}|} = |\mathbf{A}^{-1}|$
- (b) $|\mathbf{AB}| = |\mathbf{A}| |\mathbf{B}|$
- (c) $|c\mathbf{A}| = c^n |\mathbf{A}|$, where c is a scalar and n is the number of rows or columns of \mathbf{A}
- (d) If \mathbf{A} and \mathbf{B} are symmetric positive definite, then $\mathbf{A}^T = \mathbf{A}$, $\mathbf{AB} = (\mathbf{BA})^T$, $(\mathbf{A}^{-1})^T = \mathbf{A}^{-1}$, and $\mathbf{x}^T \mathbf{A}^{-1} \mathbf{y} = \mathbf{y}^T \mathbf{A}^{-1} \mathbf{x}$
- (e) If these inverses exist, and \mathbf{A} and \mathbf{B} are symmetric, then $(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{A} = \mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{B} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{B} = \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{A}$

product kernel for $\rho = 1$,

$$K_1\{f(\mathbf{x}|\boldsymbol{\mu}_o, \mathbf{P}_o), f(\mathbf{x}|\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}})\} = \frac{1}{|2\pi(\mathbf{P}_o + \hat{\mathbf{P}})|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}})^T(\mathbf{P}_o + \hat{\mathbf{P}})^{-1}(\boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}})\right\}. \quad (5.7)$$

5.2.1.2 Hellinger Distance & Hellinger Affinity Measure Closed-Form Solutions. Two closed-form approximations of the Hellinger Distance and Hellinger Affinity Measure are derived in this subsection. The first closed-form approximation uses a binomial series expansion⁴ for the $\sqrt{f(\mathbf{x}|\boldsymbol{\Omega}_o)f(\mathbf{x}|\hat{\boldsymbol{\Omega}})}$ term in Equations (4.4) and (4.6), and extracts the first-order term from this expansion as the approximation. The second approximation is the “heuristic approximation” suggested in [15] which replaces a square root of a sum of terms with the sum of the square root of each term (notionally this approximation is $\sqrt{a+b+c} \approx \sqrt{a} + \sqrt{b} + \sqrt{c}$).

To use the binomial series to expand $\sqrt{f(\mathbf{x}|\boldsymbol{\Omega}_o)f(\mathbf{x}|\hat{\boldsymbol{\Omega}})}$, first note that this term may be equivalently written as $[1 + (f(\mathbf{x}|\boldsymbol{\Omega}_o)f(\mathbf{x}|\hat{\boldsymbol{\Omega}}) - 1)]^{1/2}$. Then, the binomial series for this expression is

$$\left[1 + (f(\mathbf{x}|\boldsymbol{\Omega}_o)f(\mathbf{x}|\hat{\boldsymbol{\Omega}}) - 1)\right]^{1/2} = \sum_{u=0}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-u+1)}{u!} \left[f(\mathbf{x}|\boldsymbol{\Omega}_o)f(\mathbf{x}|\hat{\boldsymbol{\Omega}}) - 1\right]^u. \quad (5.8)$$

⁴The binomial series is given by [3]

$$(1+y)^v = \sum_{u=0}^{\infty} \frac{v(v-1)\dots(v-u+1)}{u!} y^u$$

for $v \in \mathbb{R}$ and $y \in (-1, 1]$.

The leading term is equivalently represented as

$$\begin{aligned}
\frac{\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-u+1)}{u!} &= \frac{(-1)^{u-1}}{2^u u!} 1 \cdot \underbrace{1 \cdot 3 \cdots (2u-3)}_{(u-1) \text{ terms}} \\
&= \frac{(-1)^{u-1}}{2^u u!} \underbrace{1 \cdot 3 \cdots (2u-3)}_u \cdot \frac{(2u-1)}{(2u-1)} \cdot \underbrace{\frac{2 \cdot 4 \cdots (2u-2) \cdot 2u}{2 \cdot 4 \cdots (2u-2) \cdot 2u}}_u \\
&= \frac{(-1)^{u-1}}{2^u u!} \frac{(2u)!}{2^u u!} \frac{1}{(2u-1)} \\
&= \frac{(-1)^{u-1}(2u)!}{2^{2u}(u!)^2(2u-1)}.
\end{aligned}$$

Substituting this expression into Equation (5.8) and writing out the first two terms in the series yields

$$\begin{aligned}
\left[1 + (f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) - 1)\right]^{1/2} &= 1 + \frac{1}{2} \left[f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) - 1\right] + \dots \\
&\approx \frac{1}{2} f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}). \tag{5.9}
\end{aligned}$$

This approximation uses the first-order term of the expansion and discards all other terms⁵. Finally, the approximate Hellinger Distance becomes

$$\begin{aligned}
T_H\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \sqrt{\frac{1}{2} \left\langle \sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})}, \sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})} \right\rangle} \\
&= \sqrt{1 - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}} \\
&\approx \sqrt{1 - \frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) d\mathbf{x}} \tag{5.10}
\end{aligned}$$

⁵The zeroth-order term is discarded since keeping this term would result in an indeterminant form when evaluating $\int_{-\infty}^{\infty} (1/2) d\mathbf{x}$.

and the Hellinger Affinity Measure is approximated as

$$\begin{aligned}
T_A\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \langle \sqrt{f(\mathbf{x}|\Omega_o)}, \sqrt{f(\mathbf{x}|\hat{\Omega})} \rangle \\
&= \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x} \\
&\approx \frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega}) d\mathbf{x}. \tag{5.11}
\end{aligned}$$

Notice that the integral term in both Equations (5.10) and (5.11) is the *cross-likeness* term of the ISE cost function, and it may be evaluated using Equations (5.4) and (5.7). Also, this approximation functionally represents a scaled version of the Expected Likelihood Kernel (Equation (4.3)), but conceptually it was derived from an approximation of the Hellinger Affinity Measure, which is a true distance measure.

The “heuristic approximation” is derived by approximating the integrand in Equations (4.4) and (4.6) as a sum of square roots and using the general probability product kernel in Equation (5.1) with $\rho = 1/2$. Begin by writing the term $\sqrt{f(\mathbf{x}(k)|\Omega_o(k))f(\mathbf{x}(k)|\hat{\Omega}(k))}$ as

$$\begin{aligned}
&\left\langle \sqrt{f(\mathbf{x}(k)|\Omega_o(k))}, \sqrt{f(\mathbf{x}(k)|\hat{\Omega}(k))} \right\rangle \\
&= \int_{-\infty}^{\infty} \sqrt{\sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_R(k)} p_{o,i}\hat{p}_j f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,i}, \mathbf{P}_{o,i}) f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)} d\mathbf{x}(k) \\
&\approx \sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_R(k)} \sqrt{p_{o,i}\hat{p}_j} \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,i}, \mathbf{P}_{o,i}) f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)} d\mathbf{x}(k) \\
&= \sum_{i=1}^{N_H(k)} \sum_{j=1}^{N_R(k)} \sqrt{p_{o,i}\hat{p}_j} K_{1/2}\{f(\mathbf{x}(k)|\boldsymbol{\mu}_{o,i}, \mathbf{P}_{o,i}), f(\mathbf{x}(k)|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{P}}_j)\}. \tag{5.12}
\end{aligned}$$

Next, evaluate Equation (5.1) with ρ set to one-half:

$$\begin{aligned} K_{1/2}\{f(\mathbf{x}|\boldsymbol{\mu}_o, \mathbf{P}_o), f(\mathbf{x}|\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}})\} &= (2\pi)^0 \left(\frac{1}{2}\right)^{-n/2} |\mathbf{P}_o^{-1} + \hat{\mathbf{P}}^{-1}|^{-1/2} |\mathbf{P}_o|^{-1/4} |\hat{\mathbf{P}}|^{-1/4} \\ &\quad \cdot \exp\left\{-\frac{1}{4} \left[\boldsymbol{\mu}_o^T \mathbf{P}_o^{-1} \boldsymbol{\mu}_o + \hat{\boldsymbol{\mu}}^T \hat{\mathbf{P}}^{-1} \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\dagger T} \mathbf{P}^{\dagger -1} \boldsymbol{\mu}^{\dagger}\right]\right\} \\ &= \left|\frac{1}{2}(\mathbf{P}_o + \hat{\mathbf{P}})\right|^{-1/2} |\mathbf{P}_o \hat{\mathbf{P}}|^{1/4} \exp\left\{-\frac{1}{4}(\boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}})^T (\mathbf{P}_o + \hat{\mathbf{P}})^{-1} (\boldsymbol{\mu}_o - \hat{\boldsymbol{\mu}})\right\}. \end{aligned} \quad (5.13)$$

Finally, the “heuristic approximation” to the Hellinger Distance and Hellinger Affinity Measure is found by substituting Equation (5.13) into Equation (5.12), and then substituting this result into the appropriate measure, Equation (4.4) or Equation (4.6).

5.2.2 Proposing Mixture Reduction Actions. Now that exact closed-form solutions or approximate closed-form solutions to the selected measure functions in the case of multivariate Gaussian mixture pdfs have been found, the process of proposing the reduced-component approximate target state pdf, $f(\mathbf{x}(k)|\hat{\Omega}(k))$, is outlined. Reduced-component mixture pdfs are proposed by either deleting a single component or merging two distinct components of the original full-component Gaussian mixture pdf, $f(\mathbf{x}(k)|\Omega_o(k))$, or of the approximated target state pdf from the previous mixture reduction algorithm iteration, depending on which Greedy assignment algorithm is used. At the end of each scan k , there are $N_H(k) + \{N_H(k)[N_H(k) - 1]/2\}$ possible reduced-component mixture pdfs to propose (see Section 4.2). If a single mixture component is deleted, then the proposed reduced-component pdf is the original target state Gaussian mixture pdf, but with one mixture component removed. The mixture weights of the reduced-component mixture pdf are *not re-normalized* until the MRA has reached the requisite number of components to improve the computational speed of the algorithm, which will be discussed in the next subsection. However, this improvement in run time alters the output of measure functions which calculate the length of the error vector between the original and reduced-component mixture pdfs (using the *Hilbert space* vector analogy of the previous chapter), such as

the ISE cost function and the Hellinger Distance. This issue is covered in Subsection 5.3.1. If two mixture components are merged, then the resulting approximate target state Gaussian mixture pdf is the original pdf, but with the two merged components replaced by a Gaussian mixture component with parameters specified by Equation (2.26). Re-normalization is not an issue since the equation used to calculate the merged-component mixture weight is simply the sum of the two merged components' mixture weights. So, as long as the sum of the original mixture weights equals one, the sum of the reduced-component mixture weights is also one.

5.2.3 Mating Measure Functions with Assignment Algorithms. Mating a measure function with an assignment algorithm produces an MRA. In this subsection, the four new measure function/assignment algorithm pairings listed at the end of Section 5.1 are expounded. Insights gleaned from [38] and Williams' original ISE cost-function-based MRA code are used in the development of the four new MRAs.

5.2.3.1 Correlation Measure MRA. The CM MRA is simply Williams' original MRA, but with the ISE cost function replaced by the Correlation Measure and the decision logic modified to accept measure function values closest to one instead of those closest to zero. A flowchart of the algorithm is shown in Figure 5.2. The algorithm begins by computing and storing each Correlation Measure term in Equations (5.4), (5.5), and (5.6). Parameters for all possible pairwise component mergings are calculated since they are needed when proposing reduced-component approximate Gaussian mixture pdfs when the reduction action is to merge components. Reduced-component mixture approximations to the original mixture are proposed by both deletion and merge reduction actions, and the reduced-component mixture with a Correlation Measure closest to one is declared as the optimal approximation to the original Gaussian mixture pdf (optimal in the sense of the measure used). This process continues through iterations of setting the optimal reduced-component approximate mixture pdf from the previous iteration as the “original” mixture pdf for the current iteration, until the desired number of mixture components is obtained. At the

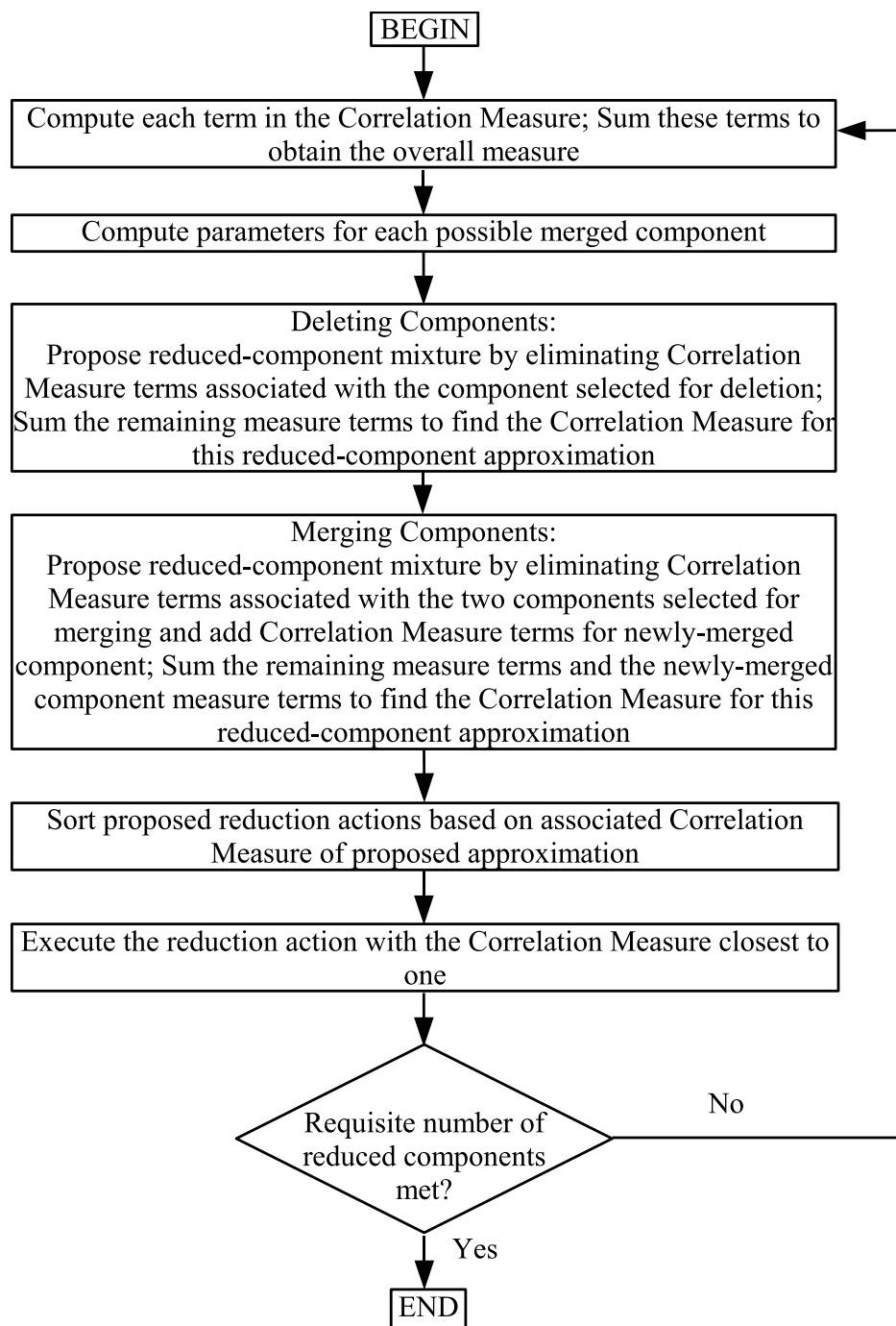


Figure 5.2: A flowchart of the Correlation Measure MRA.

final step of the algorithm, the mixture weights of the optimal approximation to the original, full-component target state Gaussian mixture pdf are re-normalized.

If the proposed reduced-component mixture is obtained by deleting a mixture component, then the corresponding Correlation Measure is found by subtracting every measure term in Equations (5.4), (5.5), and (5.6) corresponding to the deleted component. This measure calculation is accomplished more quickly if re-normalization of the remaining components' mixture weights is not considered since the Correlation Measure terms initially computed and stored at the start of the algorithm may be used. Omitting re-normalization of the mixture weights during operation of the algorithm does not degrade the approximation fidelity of MRAs based on measure functions of the cosine of the angle between two vectors representing the original and reduced-component mixtures, since re-normalization only affects the length of the vectors and not their directions. This concept is explained in detail in Subsection 5.3.1.

If the proposed reduced-component mixture is obtained by merging two mixture components, then the corresponding Correlation Measure between the full-component (or the resulting approximate mixture pdf from a previous iteration of the algorithm) and reduced-component pdfs is found by a slightly different procedure than for the case of deleting a component. First, the Correlation Measure terms containing the two merged mixture components are subtracted from the sum of the Correlation Measure terms initially calculated. Then, the Correlation Measure between the newly-formed merged component and every other surviving component is calculated according to Equations (5.4), (5.5), and (5.6). These terms are added together and then to the Correlation Measure found in the first step to obtain the Correlation Measure of the reduced-component Gaussian mixture pdf resulting from merging the two selected mixture components. Again, computational efficiency is gained by re-using previously stored measure terms as in the deletion reduction action case.

5.2.3.2 ISE Shotgun MRA. When the ISE cost function is combined with Greedy Algorithm A from Section 4.2, the ISE Shotgun MRA is created. This new MRA sacrifices the quality of the reduced-component approximation provided by Williams' original ISE cost-function-based MRA (this MRA used Greedy Algorithm B), but improves the computational speed of the algorithm. The computational acceleration may be seen in the ISE Shotgun MRA flowchart of Figure 5.3. Notice that the algorithm continues to execute the lowest-cost reduction actions based on the initially calculated costs until none remain or the requisite number of reduced components is met (as indicated by the first decision block in Figure 5.3). In contrast, Williams' original ISE cost-function-based MRA recomputes new costs after each reduction action is executed⁶.

5.2.3.3 Hellinger Distance MRA. As shown in Figure 5.4, the HD MRA is the same as the CM MRA except that the appropriate approximation to the Hellinger Distance replaces the Correlation Measure as the measure function, and the reduced-component mixture pdf with the smallest distance measure with respect to the original mixture pdf is the optimal approximation. However, neglecting mixture re-normalization causes a problem with the Hellinger Distance given by Equation (4.4). The problem lies in the simplification made in the final line of the Hellinger

⁶In [41], Williams identifies two efficiency enhancements beyond those incorporated into his original algorithm which modify the operation of his MRA by negating the need to recompute *all* of the new costs.

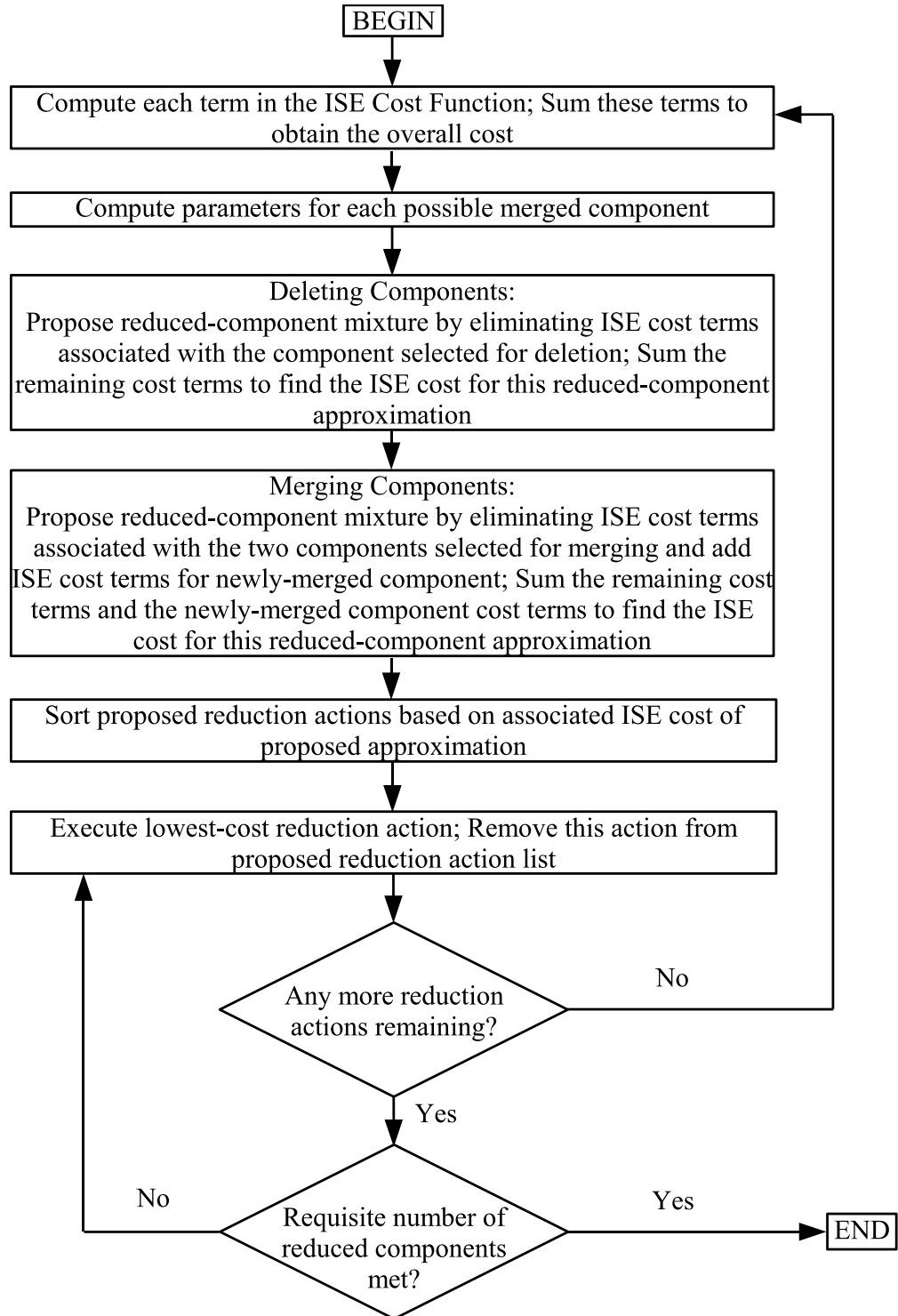


Figure 5.3: A flowchart of the ISE Shotgun MRA.

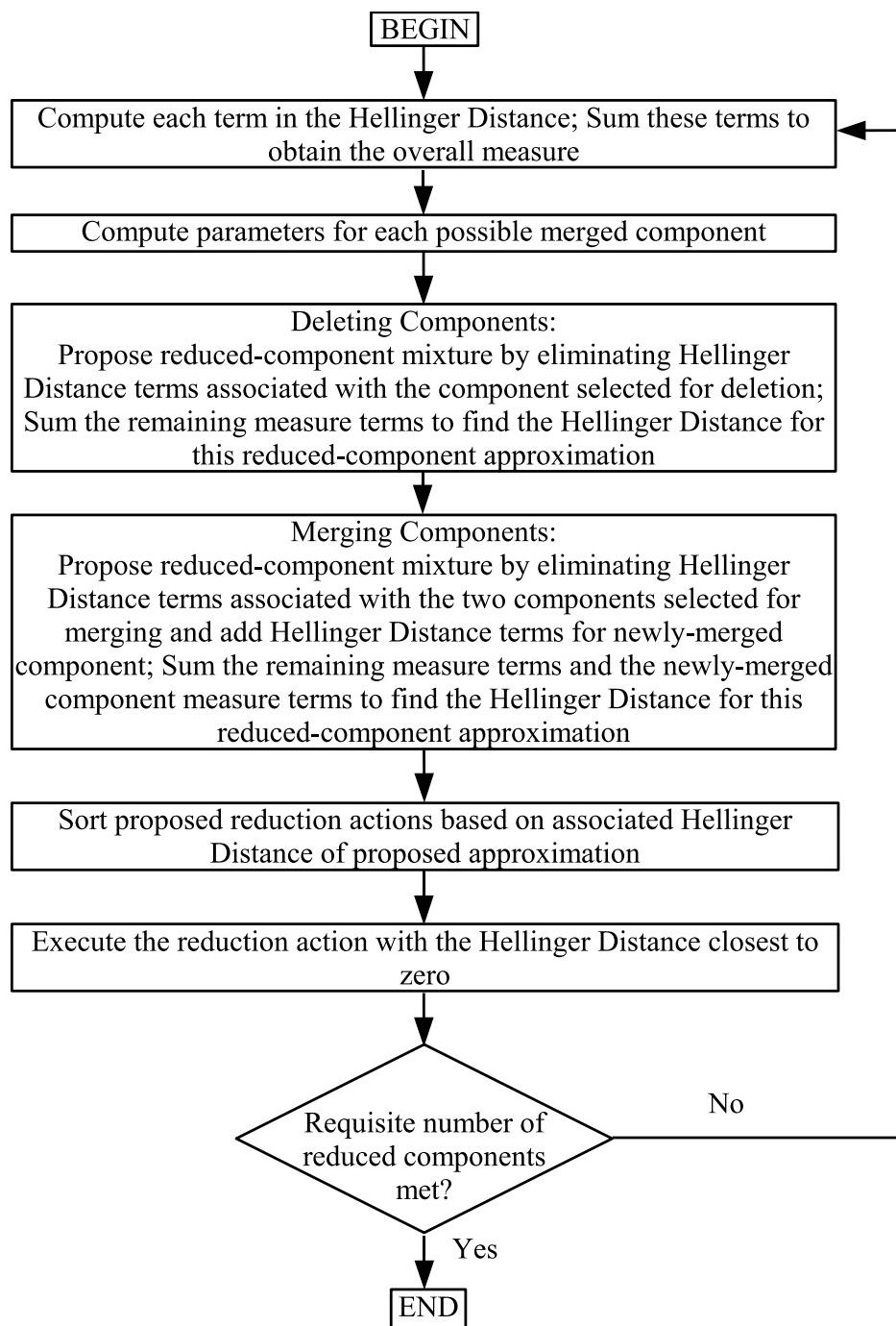


Figure 5.4: A flowchart of the Hellinger Distance MRA.

Distance equation

$$\begin{aligned}
T_H\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \sqrt{\frac{1}{2}\left\langle \sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})}, \sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})} \right\rangle} \\
&= \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} \left[\sqrt{f(\mathbf{x}|\Omega_o)} - \sqrt{f(\mathbf{x}|\hat{\Omega})} \right]^2 d\mathbf{x}} \\
&= \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o) d\mathbf{x} + \frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\hat{\Omega}) d\mathbf{x} - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}}.
\end{aligned} \tag{4.4}$$

The first integral term is one-half since $f(\mathbf{x}|\Omega_o)$ is a valid Gaussian mixture pdf with the sum of its mixture weights constrained to one. However, the second integral term is *not* guaranteed to be one during operation of the HD MRA since the reduced-component mixture weights are not re-normalized until the last step of the algorithm. Thus, the Hellinger Distance given by Equation (4.4) must be modified to include the possibility of reduced-component mixture weights which are not normalized. This modification is accomplished by using Equation (2.23) (with appropriate modification

to the mixture parameter notation):

$$\begin{aligned}
T_H\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} \\
&= \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o) d\mathbf{x} + \frac{1}{2} \int_{-\infty}^{\infty} f(\mathbf{x}|\hat{\Omega}) d\mathbf{x} - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}} \\
&= \sqrt{\frac{1}{2} + \frac{1}{2} \sum_{i=1}^M \hat{p}_i \int_{-\infty}^{\infty} f(\mathbf{x}|\hat{\mu}_i, \hat{\mathbf{P}}_i) d\mathbf{x} - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}} \\
&= \sqrt{\frac{1}{2} + \frac{1}{2} \sum_{i=1}^M \hat{p}_i - \int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}}
\end{aligned} \tag{5.14}$$

since the integral of a pdf over the entire sample space of the continuous random quantity it describes is one.

5.2.3.4 Hellinger Affinity Measure MRA. Figure 5.5 shows the flowchart for the HA MRA. The algorithm is essentially the same as the HD MRA, including a similar modification to the Hellinger Affinity Measure equation when the reduced-component mixture weights are not normalized, but special consideration is necessary when choosing the optimal value of this measure. Since the reduced-component mixture weights are not guaranteed to be normalized throughout the operation of the algorithm, the Hellinger Affinity Measure, given by Equation (4.6), is

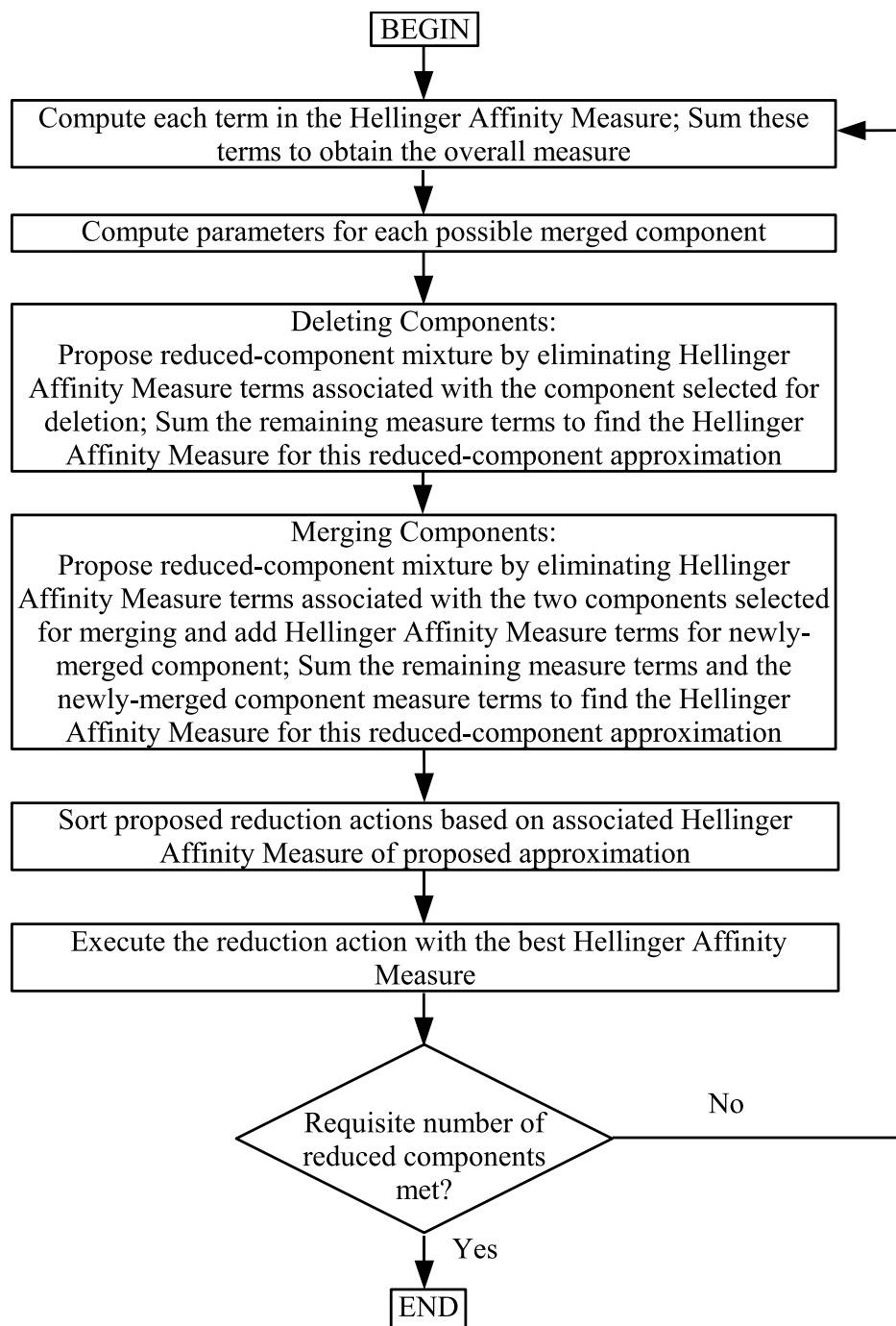


Figure 5.5: A flowchart of the Hellinger Affinity Measure MRA.

modified to account for this possibility.

$$\begin{aligned}
T_A\{f(\mathbf{x}|\Omega_o), f(\mathbf{x}|\hat{\Omega})\} &= \frac{\int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}}{\sqrt{\int_{-\infty}^{\infty} f(\mathbf{x}|\Omega_o) d\mathbf{x} \int_{-\infty}^{\infty} f(\mathbf{x}|\hat{\Omega}) d\mathbf{x}}} \\
&= \frac{\int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}}{\sqrt{\sum_{i=1}^M \hat{p}_i \int_{-\infty}^{\infty} f(\mathbf{x}|\hat{\mu}_i, \hat{\mathbf{P}}_i) d\mathbf{x}}} \\
&= \frac{\int_{-\infty}^{\infty} \sqrt{f(\mathbf{x}|\Omega_o)f(\mathbf{x}|\hat{\Omega})} d\mathbf{x}}{\sqrt{\sum_{i=1}^M \hat{p}_i}} \tag{5.15}
\end{aligned}$$

Either the truncated binomial series or the “heuristic approximation” is applied to the numerator term to find an approximate closed-form solution. However, since the quantity on the right-hand side of Equation (5.15) represents the cosine of the angle between the square root of the original and approximate pdfs using the vector analogy developed in Subsection 4.1.1, this quantity must be bounded between -1 and $+1$. If either approximation produces values outside of these bounds, then the resulting distance measure is not valid. In this case, one may choose to declare the approximation inadequate and discard it.

5.3 Mixture Reduction Algorithm Analysis

The new MRAs developed in the previous section are analyzed in this section. An analysis of the impact of neglecting mixture weight re-normalization during operation of the MRAs is presented in Subsection 5.3.1. Also, each new MRA is tested in Subsection 5.3.2 using two randomly-generated univariate Gaussian mixture pdfs,

and the approximate reduced-component mixture pdf produced by each algorithm is qualitatively compared with the original, full-component mixture and the output of the other MRAs. A comparison is also made with respect to required computation time for each algorithm when coded using the same techniques.

5.3.1 Impact of Mixture Weight Re-Normalization. Subsection 5.2.2 mentions that mixture weight re-normalization of the reduced-component approximation Gaussian mixture pdf is neglected until near the last step of each MRA resulting in improved computational speed. Using the *Hilbert space* vector analogy for Gaussian mixture pdfs of Section 4.1, this subsection shows that this improvement does not impact MRAs based on true distance measure functions of the *cosine of the angle* between two vectors representing mixture pdfs, but that, in general, it does affect MRAs based on true distance measure functions of the *error vector* between two vectors representing mixture pdfs. That is, true distance measures of the cosine of the angle between two vectors representing Gaussian mixture pdfs are invariant to scalar transformations of the mixtures, such as re-normalization of mixture weights, but true distance measures of the error vector between the two mixture pdfs are not invariant to this type of transformation. Recall that the Correlation Measure calculates the cosine of the angle between a vector representing the original mixture pdf and another vector representing the reduced-component approximation of the same, while the Hellinger Affinity Measure calculates the same quantity, but the vectors represent the square root of each pdf instead. Also, recall that the ISE cost function is the squared length of the error vector between the two vectors representing the original and approximate Gaussian mixture pdfs, and that the Hellinger Distance is the length of the error vector between two vectors representing the square root of each mixture pdf.

Figure 5.6 may be used to explain the effect of mixture weight re-normalization on the Hellinger Affinity Measure, Correlation Measure, Hellinger Distance, and ISE cost function. Imagine that a mixture component deletion action is executed and that

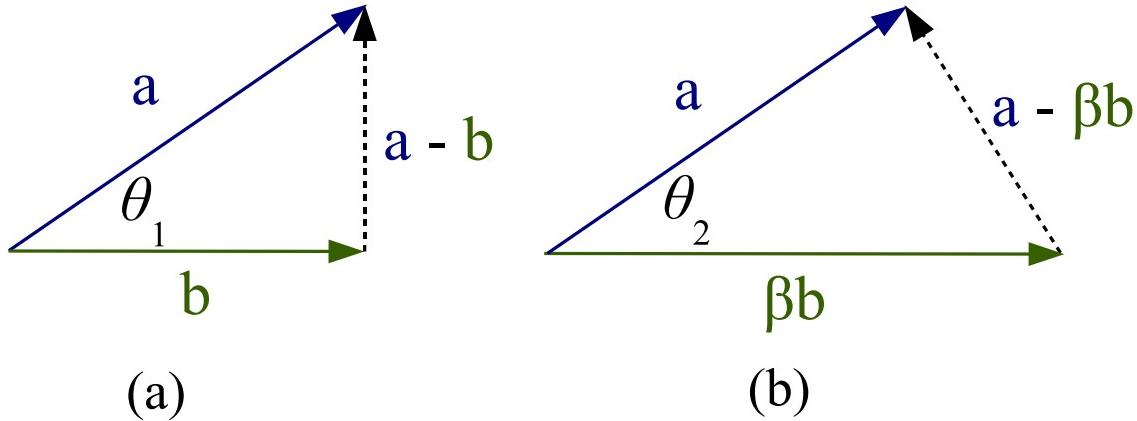


Figure 5.6: Two depictions of the error vectors and angles between the vectors a , b and βb which represent the original mixture pdf, a reduced-component approximation mixture pdf, and a scaled (re-normalized) version of the same reduced-component approximation pdf, respectively. Notice that the length of the error vector $a - b$ is different from that of the error vector $a - \beta b$, but that the angles θ_1 and θ_2 are the same.

the mixture weights of the approximation pdf are *not* re-normalized so that the vector b , as shown in sub-plot (a), represents the reduced-component approximation. Also, consider the case in which mixture weight re-normalization is performed so that the approximation pdf is represented by the scaled vector βb , as shown in sub-plot (b). Notice in sub-plots (a) and (b) that the angles θ_1 and θ_2 are equal, but that the lengths of the error vectors $a - b$ and $a - \beta b$ are different. Since the Correlation Measure is the cosine of the angle between two vectors, mixture weight re-normalization has no effect on this measure function. Likewise, mixture weight re-normalization has no impact on the Hellinger Affinity Measure as long as the appropriate modification, given by Equation (5.15), is made to this measure function. In contrast, the ISE cost function and the Hellinger Distance are dependent on the length of the vector representing the reduced-component approximation mixture pdf, so, in general, re-normalization of the approximation pdf mixture weights affects the output of these measure functions.

As an illustration of the role re-normalization plays in an MRA, consider reducing a 19-component univariate Gaussian mixture pdf to one containing only eight components using the ISE and CM MRAs with and without mixture weight re-

normalization⁷. Figure 5.7 shows the outputs of the ISE and CM MRAs when re-normalization of the reduced-component approximation mixture pdf is *not* performed (sub-plots (a) and (b)), and when re-normalization is performed (sub-plots (c) and (d)). As predicted by theory, the CM MRA produces the same result regardless of whether or not mixture weight re-normalization is performed during operation of the algorithm, since the output of the Correlation Measure for the reduced-component approximation pdf and the re-normalized (scaled) version of the approximation pdf is the same. However, a comparison of sub-plots (a) and (c) shows that the ISE MRA is dependent on mixture weight re-normalization, since the ISE cost function depends on the length of the reduced-component approximation mixture pdf.

Although in general the ISE and HD MRAs are dependent on mixture weight re-normalization during their operation, multiple experiments using the ISE and HD MRAs have shown that this dependence is usually not strongly evident. As an example, if the 19-component univariate Gaussian mixture pdf from the previous paragraph is reduced to a four-component mixture pdf instead of an eight-component one, then the reduced-component approximation Gaussian mixture pdf is the same whether or not mixture weight re-normalization is performed (however, the intermediate steps of the reduction process differ between the two implementations). Considering the substantial computational savings of *not* re-normalizing the mixture weights until near the last step of an MRA and the unlikelihood of encountering a reduction task which would result in significantly different outputs depending on whether or not mixture weight re-normalization is performed, neglecting re-normalization until near the last step of an MRA appears to be worth any potential degradation in approximation quality. Therefore, all of the MRAs tested in the next subsection do not re-normalize the mixture weights of the reduced-component approximation pdf until near the end of the algorithm.

⁷The HD MRA and HA MRA are not considered since evaluation of their corresponding measure functions requires approximation which could potentially obscure the effect of mixture weight re-normalization.

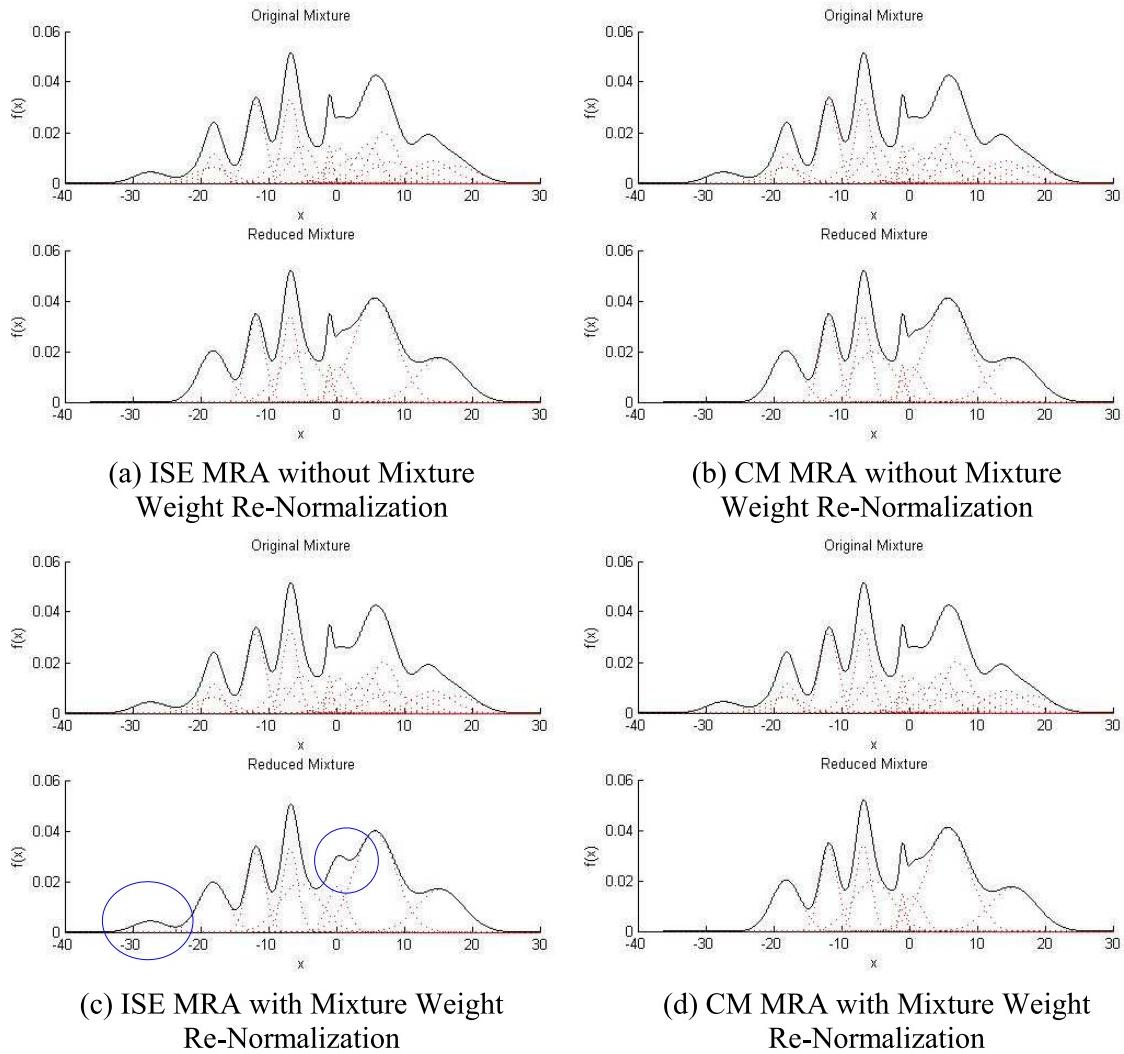


Figure 5.7: An illustration of the effect of re-normalizing the mixture weights of a reduced-component approximation Gaussian mixture pdf using the ISE and CM MRAs with and without re-normalization during operation of the algorithms. Notice that re-normalizing the mixture weights of the approximation pdf has no effect on the CM MRA output, but the ISE MRA produces different results (the circled portions of the pdf in subplot (c)) depending on whether mixture weight re-normalization is used or not.

5.3.2 Mixture Reduction Algorithm Test Results. Two randomly-generated univariate Gaussian mixture pdfs are used to test the fidelity of the reduced-component approximation Gaussian mixture pdfs produced by the new MRAs and Williams' original ISE MRA. MRAs are judged based on the aesthetic quality of their respective approximations to the full-component Gaussian mixture pdf and on the required computation time for each algorithm when coded using the same techniques. During testing, it was noted that the "heuristic approximation" for the Hellinger Distance and Hellinger Affinity Measure produced invalid distance measures. Computed distance measures for the Hellinger Distance were negative, and the Hellinger Affinity Measure produced outputs which were greater than +1. Since the resulting distance measures were invalid, the HD and HA MRAs based on this approximation were abandoned, and only the HD and HA MRAs based on the truncated binomial series approximation were considered.

The ISE, CM, ISE Shotgun, HA, and HD MRAs were tested against two randomly-generated univariate Gaussian mixture pdfs. Results for the first test, which was to reduce a 15-component mixture to a 10-component one, are shown in Figure 5.8. Of the five MRAs tested, the ISE Shotgun MRA produced the worst-looking approximation, and the ISE and CM MRAs generated reduced mixture approximations that appear almost identical to the original mixture pdf. The HD and HA MRAs produced questionable approximations, but required about half of the computation time as that of the ISE and CM MRAs. Almost the same reduction in computation time was noted for the ISE Shotgun MRA as well.

Figure 5.9 displays the results of the second MRA test, which required each MRA to reduce a 19-component univariate Gaussian mixture pdf to a 5-component mixture. Again, the ISE and CM MRAs produced the best-looking approximations despite reducing the number of components in the original mixture pdf by almost 75%. However, in contrast to the first test, the ISE Shotgun MRA generated a visibly better reduced-component approximation than either of the Hellinger-based MRAs using the truncated binomial series approximation. As in the first test, the ISE Shotgun, HD,

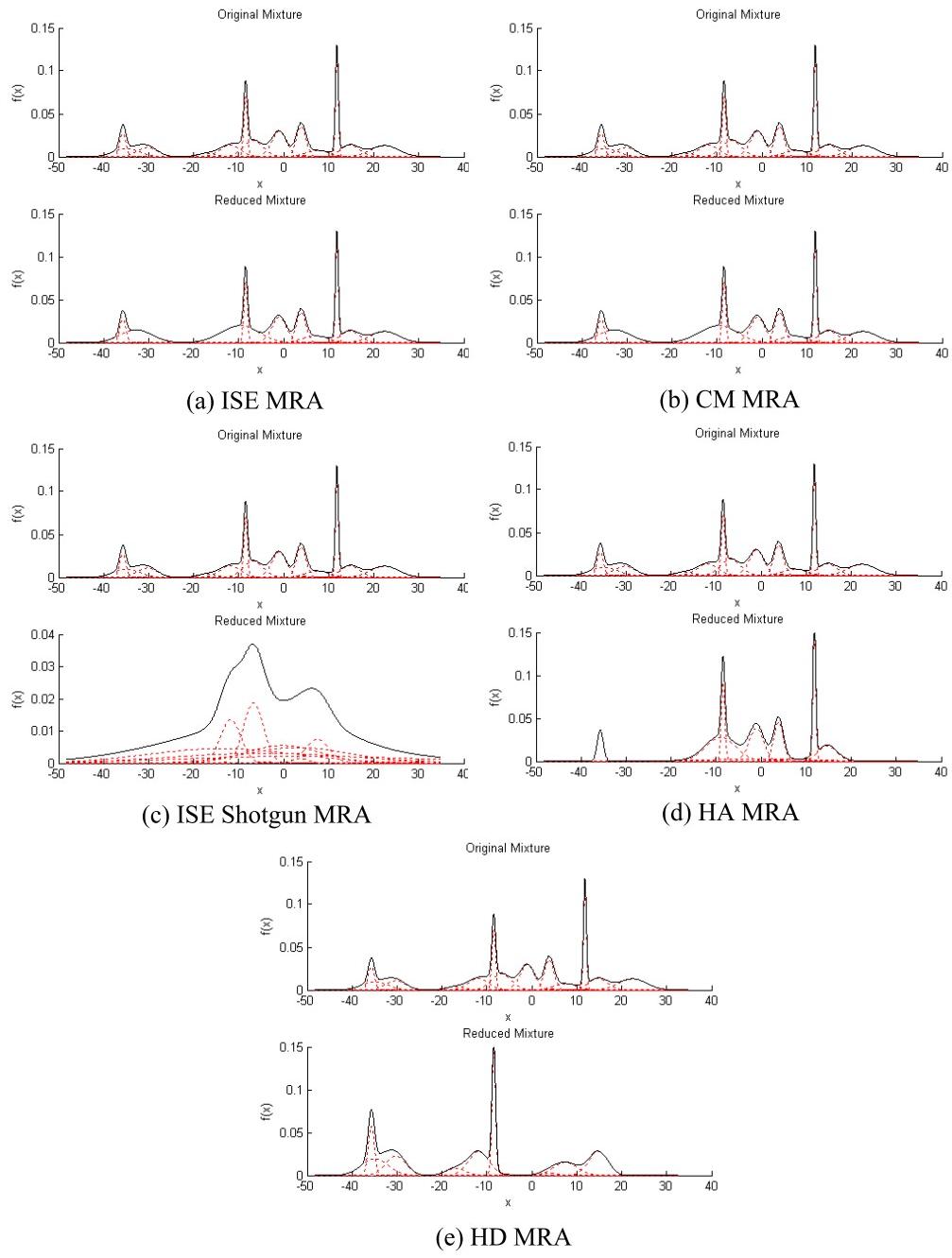


Figure 5.8: The first MRA test consisting of a 15-component univariate Gaussian mixture pdf which is approximated by a 10-component mixture pdf using the corresponding MRA. Mixture components are represented by dashed traces while the complete mixture pdfs are depicted by the solid traces.

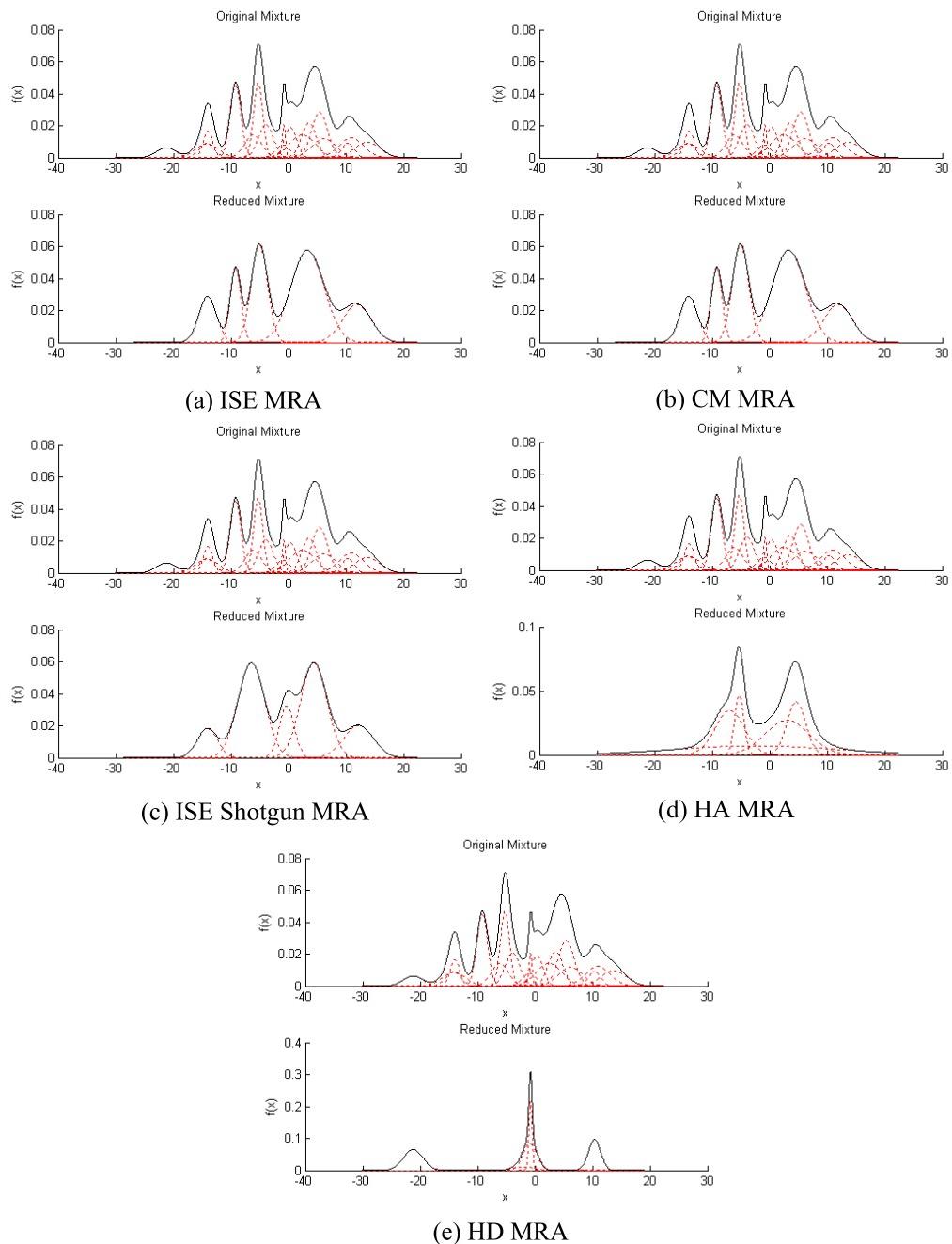


Figure 5.9: The second MRA test consisting of a 19-component univariate Gaussian mixture pdf which is approximated by a 5-component mixture pdf using the corresponding MRA. Mixture components are represented by dashed traces while the complete mixture pdfs are depicted by the solid traces.

and HA MRAs took almost half of the required computation time as that for the ISE and CM MRAs.

Based on the qualitative assessments of the approximations produced by the MRAs in the two tests, the CM MRA is the best new candidate MRA (in addition to William's ISE MRA) for use in a Bayesian tracking in clutter algorithm. The HA and HD MRAs performed rather poorly, which is likely attributable to the crude approximation used in Equation (5.9). Utilizing higher-order terms in the binomial series of Equation (5.8) would likely improve performance. However, the added computational requirements may outweigh any performance gains, especially if the additional computations necessary for the *approximation* lead to a similar computational requirement as that for the ISE and CM MRAs, which use *exact* closed-form solutions for their respective measure functions. Results for the ISE Shotgun MRA do not indicate that this MRA is a good candidate for implementation into a Bayesian tracking in clutter algorithm: Greedy Algorithm B is significantly superior to Greedy Algorithm A for MRA performance.

5.4 Summary

Four new mixture reduction algorithms (MRAs) were developed, implemented, and tested in this chapter. MRAs based on the Hellinger Distance and Hellinger Affinity Measure used either the truncated binomial series approximation or the “heuristic approximation” suggested in [15]. However, the second approximation produced invalid distance measures, so it was abandoned in favor of the truncated binomial series approximation. Williams’ original Integral Square Error (ISE) cost-function-based MRA was modified by replacing Greedy Algorithm B with Greedy Algorithm A to create the ISE Shotgun MRA, and it was also modified by swapping the ISE cost function with the Correlation Measure (CM) to produce the CM MRA. Of the four new MRAs proposed in this chapter, only the CM MRA appears suitable for use in a Bayesian tracking in clutter algorithm: only it yields performance comparable to that of the ISE MRA.

VI. Simulation Description & Results

In the previous chapter, the Correlation Measure mixture reduction algorithm (CM MRA) was selected as the best candidate as an alternative to the Integral Square Error (ISE) MRA for the mixture reduction portion of a practical Bayesian tracking in the presence of measurement origin uncertainty algorithm. The single-target in heavy clutter simulation scenario presented in [38] is used to test both the CM MRA and Williams' ISE cost-function-based MRA. Simulation results for each MRA are compared, and a final evaluation of the relative performance of the CM MRA is presented.

6.1 Description

The single-target in heavy clutter scenario found in [38] is used to test both the CM MRA and the ISE MRA. The target travels in the x - y plane according to the constant velocity (CV) model of Section 2.1,

$$\mathbf{x}(k) = \begin{bmatrix} x(k) \\ v_x(k) \\ y(k) \\ v_y(k) \end{bmatrix} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ v_x(k-1) \\ y(k-1) \\ v_y(k-1) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{bmatrix} \begin{bmatrix} w_x(k-1) \\ w_y(k-1) \end{bmatrix} \quad (2.8)$$

with $E\{w_x(k)\} = E\{w_y(k)\} = 0$, $E\{w_x(k)w_x(l)\} = \delta_{kl}$, $E\{w_y(k)w_y(l)\} = \delta_{kl}$, $E\{w_x(k)w_y(l)\} = 0$, and $T = 1$ second. Noise-corrupted measurements of the x and y positions are available at each sample k , and the measurement model is

$$\mathbf{z}(k) = \begin{bmatrix} z_x(k) \\ z_y(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ v_x(k) \\ y(k) \\ v_y(k) \end{bmatrix} + \begin{bmatrix} n_x(k) \\ n_y(k) \end{bmatrix} \quad (6.1)$$

where $E\{n_x(k)\} = E\{n_y(k)\} = 0$, $E\{n_x(k)n_x(l)\} = \delta_{kl}$, $E\{n_y(k)n_y(l)\} = \delta_{kl}$, and $E\{n_x(k)n_y(l)\} = 0$. Initial conditions for the Gaussian target state random process

vector are

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} 0 \\ 10 \\ 0 \\ 10 \end{bmatrix}, \quad \mathbf{P}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6.2)$$

At each scan k , the expected number of false-origin measurements, $\lambda_{FT}V_S$, is 480. The false-origin measurement clutter density, λ_{FT} , is set to 0.012 so that $V_S = 200 \times 200$, which is the surveillance region of the sensor at scan k . This region is a box in the x - y plane centered at the true target location at sample k . The target-oriented data association approach and measurement gating are utilized, and the gate probability, P_g , is set to one (P_g is the probability that the true target measurement falls within the corresponding measurement gate). Measurement gating is accomplished using Williams' square gating routine in which the gate is formed as the square centered about a predicted measurement with side length of twice the square root of the maximum eigenvalue of the covariance of the residual after scaling by the gate threshold [38, 40, 41]. Finally, the probability of detection, P_d , is also set to one.

Two-hundred Monte Carlo simulations were run using Williams' MATLAB® code [38, 40, 41] with the maximum number of mixture components set to 1, 5, 10, 15, 20, 25, 30, and 35 components using both the ISE and CM MRAs. The CM MRA was implemented by replacing the ISE cost function with the Correlation Measure in Williams' ISE MRA MEX C-code and modifying the decisions criterion accordingly. Each simulation determined the total number of scans that the target was tracked before track loss occurred (i.e., the track life). Track loss occurs if either of the following criteria are met:

- (i) the true target measurement is not within the measurement gates of any of the hypothesized tracks for five consecutive scans, or

- (ii) the discrepancy between the combined target state mean estimate for every hypothesized track and the combined true target state is greater than ten standard deviations¹ for five consecutive scans.

Pseudo-random number generators were set to predetermined values at the beginning of each simulation so that both MRAs were presented with exactly the same measurement data before track loss occurred. The track life results of these simulations are presented in the next section.

6.2 Results

The results from the simulation scenario described in the last section are covered in this section. Figure 6.1 depicts the average track life for the CM and ISE MRAs as a function of the maximum number of mixture components over each set of two-hundred Monte Carlo trials. As expected, the average track life improves as the number of mixture components increases, since including more components in the target state pdf approximation produces a better representation of the original target state Gaussian mixture pdf. The CM MRA appears to outperform the ISE MRA slightly in some cases while the opposite is true in other cases, however, overall the average track life differences between the two MRAs are statistically insignificant.

Figure 6.2 shows the percentage and number of individual trials in which the track life of the CM MRA was *exactly* the same, better than, and worse than the track life of the ISE MRA. For instance, the bar on the far left of the figure (for the case in which the maximum number of mixture components is set to one) shows that a large percentage (99%) of the trials resulted in exactly the same track life for both algorithms and that in two of those trials the MRAs produced different track life results. In one of those trials in which the results of the MRAs differed, the CM MRA produced a longer track life than the ISE MRA, and the ISE MRA generated a better track life in the other trial. Although it may seem somewhat surprising that

¹As calculated by a Kalman Filter applied to the same simulation scenario in the absence of clutter [31, 38].

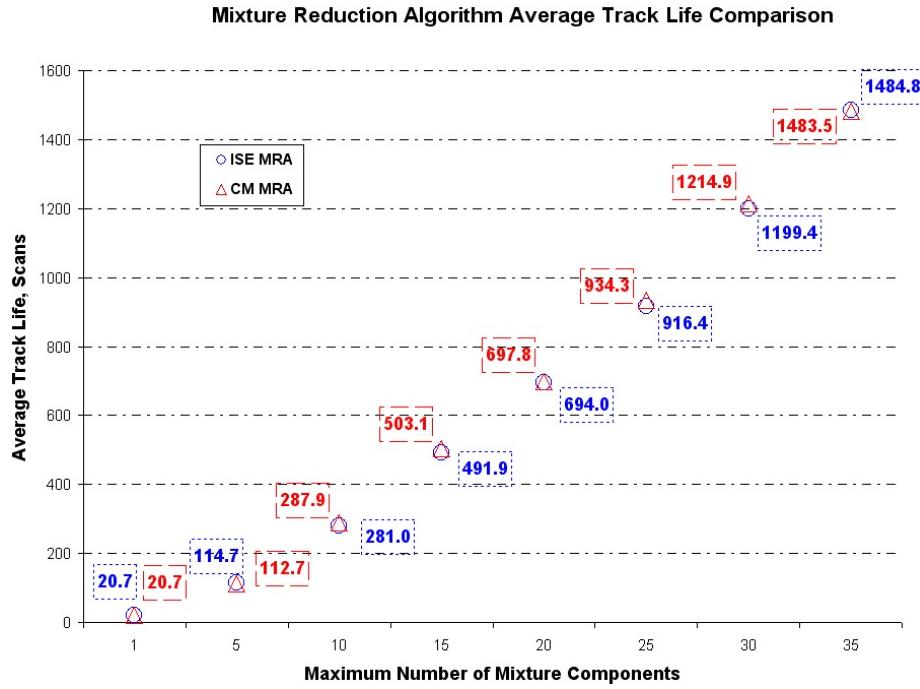


Figure 6.1: Average track life results for the CM and ISE MRAs. The figures within the coarsely-dashed boxes correspond to CM MRA results, while the numbers inside of the finely-dashed boxes are for the ISE MRA.

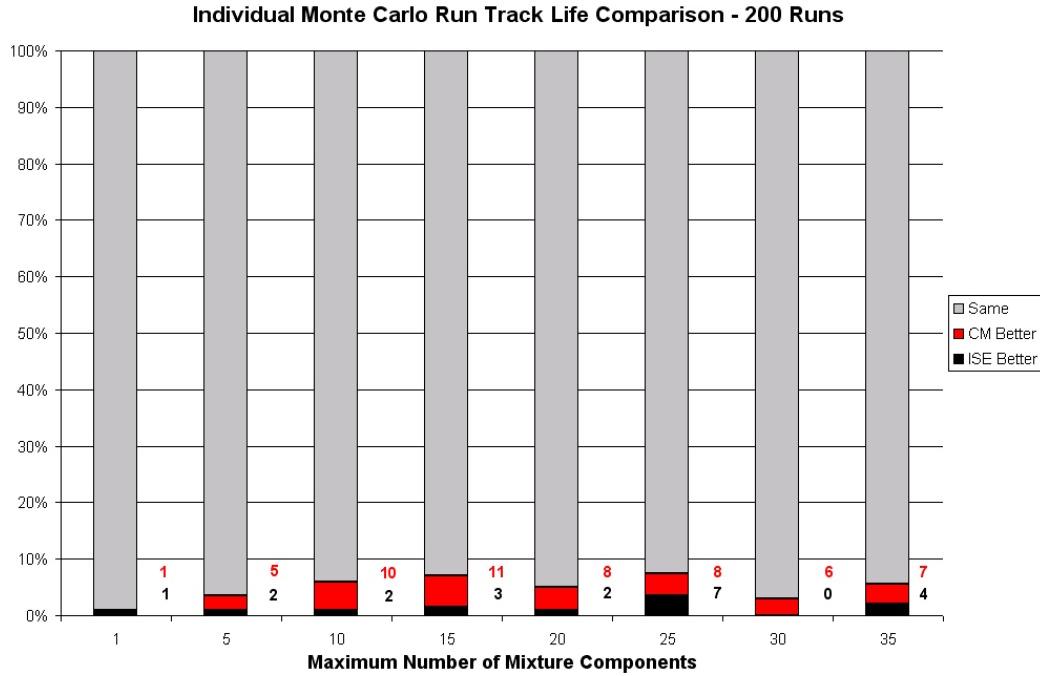


Figure 6.2: The percentage and number of trials in which the track life of the CM MRA was exactly the same, better than, and worse than that of the ISE MRA.

the majority of the trials resulted in exactly the same track life figures for the CM and ISE MRAs, recall that these MRAs produced the same reduced-component univariate Gaussian mixture pdfs, as shown in Figures 5.8 and 5.9, for the two tests conducted in Subsection 5.3.2, so this outcome should not be unexpected. Notice that in cases in which the track life results of the two MRAs differ, the CM MRA outperforms the ISE MRA for most settings of the maximum number of mixture components. However, this graphic does not indicate the number of scans by which the CM MRA track life is longer for these cases. This information is contained in the next figure.

For cases in which the track life for the CM and ISE MRAs differ, Figure 6.3 shows the average, maximum, and minimum track life disparities between the two MRAs². The upper plot, (a), shows that, for four of the seven settings for the maximum allowable number of mixture components, in cases in which the CM MRA outperformed the ISE MRA, the average disparity between the track life of the two MRAs is larger than that for cases in which the ISE MRA outperformed the CM MRA. In five of the seven settings for the maximum number of mixture components, the maximum difference between the two MRAs is greatest for the CM MRA in trials in which the CM MRA outperformed the ISE MRA. Sub-plot (b) is an enlarged version of sub-plot (a). This graphic shows values in which the track life disparity between the CM and ISE MRAs is less than ten scans. Notice that track life difference metrics are not shown for the ISE MRA when the maximum allowable number of mixture components is set to 30 since, as shown in Figure 6.2, the ISE MRA did not outperform the CM MRA in any of the trials.

²The single mixture component simulations were not included in this figure to minimize clutter. In the two trials in which the CM and ISE MRAs differed in track life, the CM MRA outperformed the ISE MRA by 22 scans in the first instance, while the ISE MRA outperformed the CM MRA by 21 scans in the other trial.

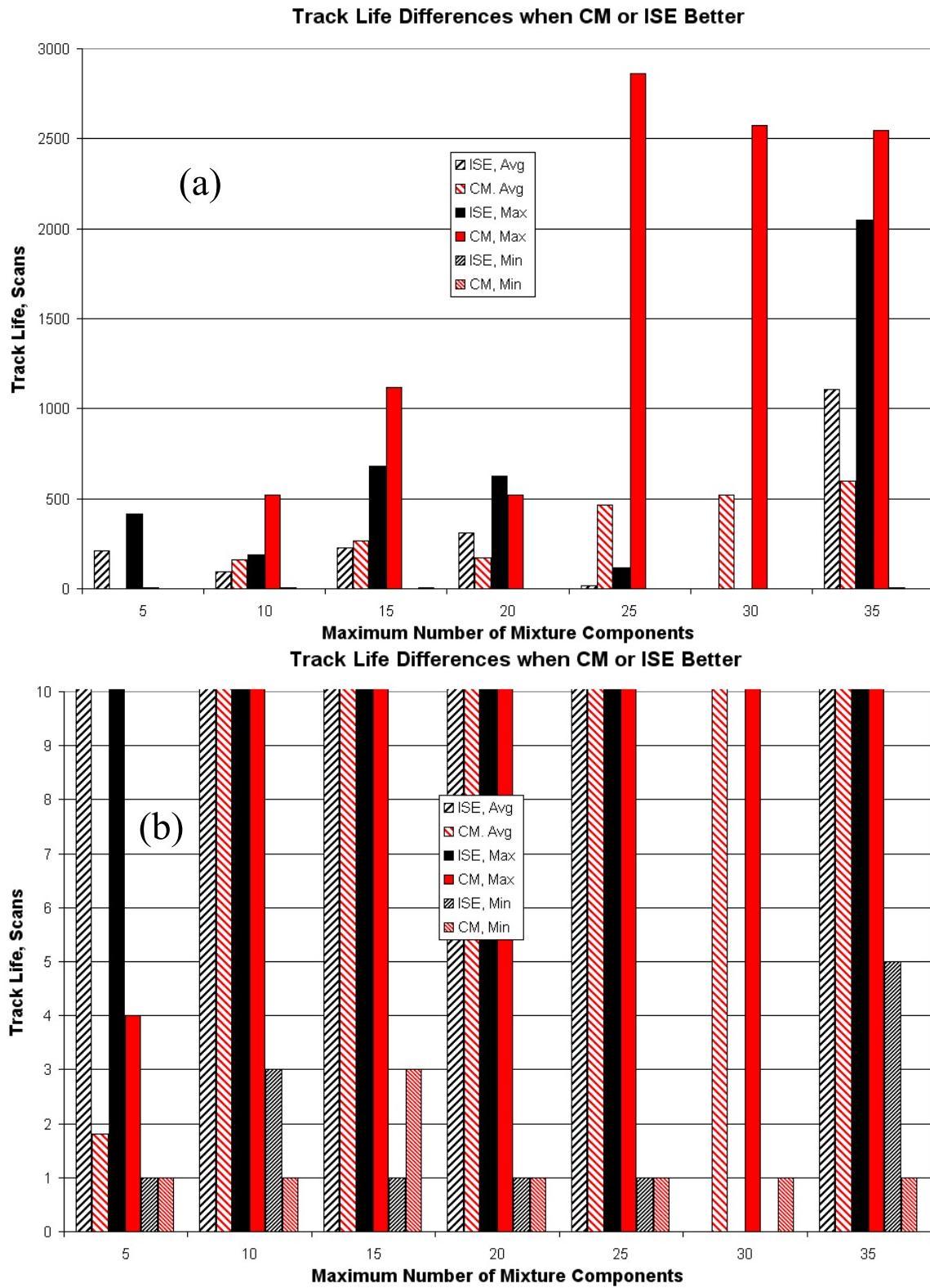


Figure 6.3: The average, maximum, and minimum track life differences between the CM and ISE MRAs in the trials in which the MRAs produced different track life results (sub-plot (b) is an enlarged version of sub-plot (a)).

6.3 Summary

This chapter presented the single-target in heavy clutter simulation scenario used to test the track life performance of the Correlation Measure mixture reduction algorithm (CM MRA) when implemented as the mixture reduction mechanism of a Bayesian tracking in clutter algorithm, in comparison to Williams' Integral Square Error (ISE) MRA. Various metrics of the track life results were obtained for the CM MRA and compared to those for the ISE MRA using the same simulation scenario. The average track life differences between the two MRAs are statistically insignificant. Since the CM and ISE MRAs only differ in the measure function used in each algorithm and each measure function requires the evaluation of all the same component terms, computation time for both MRAs is almost exactly the same under the condition that both MRAs produce the same track life³. Trials in which the track life figures for the two MRAs differed were further analyzed, but a conclusive declaration about the superiority of one MRA over the other could not be made based on the test data.

³In general, the scalar multiplication and division operations used to compute the Correlation Measure are more computationally costly than the scalar addition and subtraction operations used in the ISE cost function. However, it was noted in various simulations that the run time for the ISE MRA and CM MRA were virtually identical.

VII. Conclusions & Recommendations

This chapter concludes the thesis by summarizing its important findings. A restatement of the research goal is presented to reacquaint the reader with the objective of the thesis, followed by a summary of key results. Significant contributions contained in this thesis are highlighted, and recommendations for future research are included in the final section.

7.1 Restatement of the Research Goal

Equation (2.65) (modified for a single target state random process vector),

$$f(\mathbf{x}(k), \Theta(k) | \mathbf{Z}^k) = \sum_{i_k \in N_H(k)} \cdots \sum_{i_1 \in N_H(1)} p(\{\Theta_{i_\ell}\}_1^k | \mathbf{Z}^k) f(\mathbf{x}(k) | \{\Theta_{i_\ell}\}_1^k, \mathbf{Z}^k)$$

of Section 2.5 is the Bayesian solution for tracking a target in clutter. As new measurements are received at subsequent scans, new summations are added, and the number of terms needed to evaluate the target state multivariate Gaussian mixture pdf becomes computationally unrealistic. Thus, some type of approximation is necessary to implement the rigorous Bayesian solution for the target state pdf.

A mixture reduction algorithm (MRA) is one method of approximation. When tracking a single target in heavy clutter while retaining a large number of mixture components, Williams' Integral Square Error (ISE) MRA has been shown to provide longer average track life results than any other algorithm [38, 40, 41]. However, recommendations in [38] indicate the potential for improving upon the results produced by Williams' algorithm. Thus, *the goal of this research is to create a new MRA which offers better tracking performance and/or decreased computation time as compared to Williams' ISE MRA.*

7.2 Summary of Results

Four new MRAs were developed in an attempt to meet the research goal. The motivation for three of the four new MRAs, the Integral Square Error (ISE) Shot-

gun, Hellinger Distance (HD), and Hellinger Affinity Measure (HA) MRAs, was the prospect of decreased computation time while preserving comparable tracking performance to the ISE MRA. However, as any engineer knows, improved performance in one aspect of a design often leads to decreased performance in another area, and the ISE Shotgun, HD, and HA MRAs were no exception. Although all three MRAs required roughly half of the computation time of Williams' original ISE MRA¹, the quality of the reduced-component univariate Gaussian mixture pdf approximations produced by these MRAs was rather poor, as shown in Figures 5.8 and 5.9.

The fourth new MRA, the Correlation Measure (CM) MRA, provided much better results than the other new MRAs, but it only offered a slight improvement over Williams' ISE MRA. Figures 5.8(b) and 5.9(b) clearly show that the reduced-component mixture pdf approximations generated by the CM MRA closely match the original univariate Gaussian mixture pdf. In fact, the CM MRA made the *same* approximations as the ISE MRA. Although this fact is not true in general (as indicated by the track life performance of the two MRAs), this phenomenon likely occurred in over ninety percent of the simulation trials, which resulted in *exactly* the same track life for both MRAs, as depicted in Figure 6.2.

Geometrically, the similarity inferred from the descriptions of each measure function given in Subsection 4.1.1 between the Correlation Measure and the ISE cost function may be used to explain this phenomenon. Using a vector analogy, the original Gaussian mixture pdf and reduced-component approximation may be thought of as two vectors a and b , respectively, in *Hilbert space*, as shown in Figure 7.1. The Correlation Measure is the *cosine of the angle*, θ , between the two mixture pdfs, and the ISE cost function is the squared length of the error vector, $(a - b)$, using the “standard” definition of the Euclidean norm. Based on the Correlation Measure, two mixtures are perfectly matched if the angle between them is zero. However, two mix-

¹This statement only pertains to MRAs implemented for the purpose of this thesis. These algorithms were coded using the same blocks of code whenever possible to minimize the impact of specific implementations on algorithm run times.

tures with a zero-angle between them may not be perfectly matched in the sense of $(a - b)$ being zero, as shown in Figure 7.1(b). In this case, the ISE cost function provides a better measure of the similarity between the two mixtures since it would indicate that the two mixtures in Figure 7.1(b) are *not* perfectly matched (the error vector is non-zero). This scenario explains a situation in which the Correlation Measure and ISE cost function may produce different results, which would explain the disparity in track life results experienced in certain simulation outcomes. In contrast, Figure 7.1(c) shows the only scenario in which the ISE cost function and Correlation Measure would “agree” on the similarity between the two mixtures, since both the angle and error vector are zero. This case may explain why over ninety percent of the simulation trials produced exactly the same track life. However, this conclusion implies that the vectors representing the full-component and reduced-component mixture pdfs are essentially co-linear and essentially of the same length over ninety percent of the time, which seems unlikely. So this explanation of the reason why ninety-plus percent of the simulation trials produced exactly the same track life is not completely satisfying.

Despite the “zero angle, non-zero error vector” deficiency of the Correlation Measure pointed out in the previous paragraph, the CM MRA slightly outperformed the ISE MRA in some aspects for the tracking scenario considered in the simulations. The CM MRA produced average track life figures for six of the eight settings of the maximum number of reduced mixture components that were as good as, or slightly better than, those for the ISE MRA, as shown in Figure 6.1. In cases in which the track life figures of the two MRAs differed, the CM MRA had more trials with longer track life values than the ISE MRA for seven of the eight settings of the maximum number of reduced mixture components (see Figure 6.2). Additionally, other metrics, depicted in Figure 6.3, indicate that the CM MRA outperformed the ISE MRA by a small margin. However, for all cases in which the CM MRA slightly outperformed the ISE MRA, the small differences in performance are statistically insignificant. In addition, computational requirements for both MRAs are effectively the same since

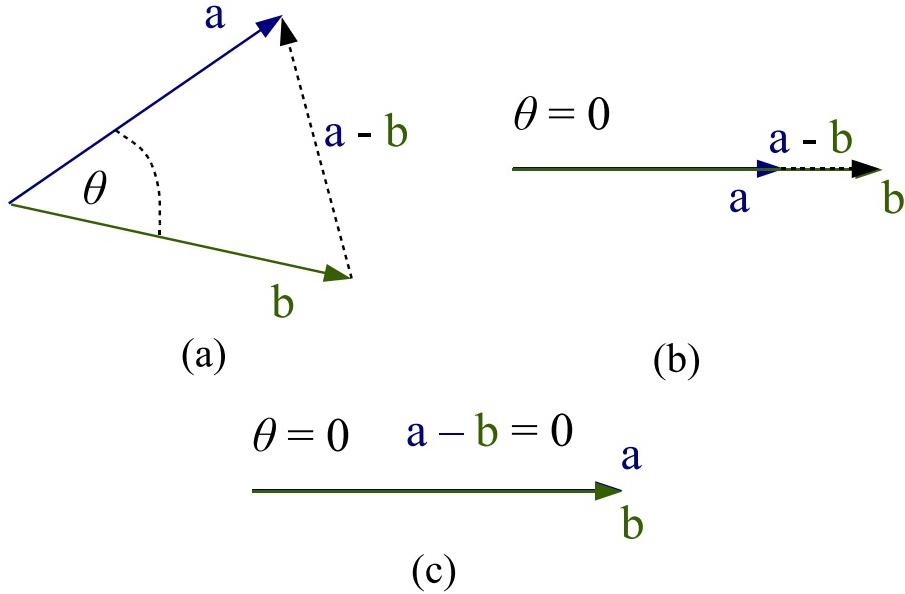


Figure 7.1: (a) Two vectors with a non-zero angle and a non-zero error vector. (b) Two vectors with a zero angle and a non-zero error vector. (c) Two vectors with a zero angle and a zero error vector.

the measure functions for both MRAs share the same components. Therefore, based on the simulation results, a vote in favor of one MRA over the other cannot objectively be made on the basis of resulting tracker performance.

7.3 Significant Contributions of Research

Although the CM MRA did not decisively outperform the ISE MRA, the CM MRA provides a viable and readily-implemented alternative for Bayesian tracking algorithms incorporating Williams' MRA. The CM MRA was shown to provide slightly better performance in certain aspects than the ISE MRA for the given simulation scenario described in Section 6.1, and it may also provide slightly better performance in other scenarios (although there is no guarantee to this claim). Additionally, the CM MRA may be fully implemented without requiring mixture weight re-normalization during operation of the algorithm, as noted in Subsection 5.3.1, thus drastically reducing the run time of the MRA. Since all of the components of the ISE cost function (the two *self-likeness* terms and one *cross-likeness* term), given by Equation (4.2),

are also used in the Correlation Measure, given in Equation (4.5), one may readily replace one measure function with the other. This interchangeability of the two measure functions provides a designer with an added degree of freedom.

7.4 Recommendations for Future Research

Several future research topics may be spawned from the research presented in this thesis. First, a hybrid Greedy Algorithm A/Greedy Algorithm B assignment algorithm, mated with either the ISE cost function or Correlation Measure, could be developed which operates according to a measure threshold. This new MRA would combine the iterative operation of Greedy Algorithm B with the computational efficiency of Greedy Algorithm A to execute all reduction actions iteratively with computed measures which do not exceed some threshold. Essentially, it would operate in a similar manner as the ISE Shotgun MRA, but instead of executing all reduction actions in one step, the algorithm would only execute those reduction actions with measures that met the threshold criteria (similarly to Salmond's Joining algorithm). For example, if the ISE cost function is used, then the algorithm would iteratively execute reduction actions which have costs below some pre-specified threshold. Thereafter, one could use the Greedy Algorithm B to drive to the final desired number of components in the reduced mixture. This new MRA should decrease computation time while, possibly, not suffering as much from the poor mixture approximation performance of the ISE Shotgun MRA.

A second potential topic for future research is using the EM algorithm of Chapter III to generate a reduced-component Gaussian mixture pdf from a full-component one. Equation (3.24) could be used as a sample-based approach to generating the reduced-component target state pdf approximation to the original target state pdf. This topic would require a radical departure from the MRAs described in this thesis, and, as a result, is a riskier potential research topic than the first one.

Finally, new approximations could possibly be developed for those true and pseudo-distance measure functions listed in Section 4.1 which do not have exact closed-

form solutions. New approximations for the Hellinger Distance and Hellinger Affinity Measure could be made by using a different series approximation than the truncated binomial series developed in Subsection 5.2.1.2. Also, as suggested in correspondence with Williams [39], a closed-form approximation to the Kullback-Leibler Mean Information and Kullback-Leibler Divergence could be developed. These pseudo-distance measure functions are popular in the literature, and an adequate approximation may even already exist.

Bibliography

1. Alspach, Daniel L. "A Gaussian Sum Approach to the Multitarget Identification–Tracking Problem". *Automatica*, 11(3):285–296, May 1975.
2. Alspach, Daniel L. and Harold W. Sorenson. "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations". *IEEE Transactions on Automatic Control*, AC-17(4):439–448, August 1972.
3. Arfken, George B. and Hans J. Weber. *Mathematical Methods for Physicists*. Harcourt Academic Press, New York, NY, fifth edition, 2001.
4. Bar-Shalom, Yaakov and Xiao-Rong Li. *Estimation and Tracking: Principles, Techniques and Software*. Artech House, Norwood, MA, 1993.
5. Bar-Shalom, Yaakov and Xiao-Rong Li. *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, Storrs, CT, 1995.
6. Blackman, Samuel S. *Multiple-Target Tracking with Radar Applications*. Artech House, Norwood, MA, 1986.
7. Blackman, Samuel S. and Robert Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, Boston, MA, 1999.
8. Brookes, M. "The Matrix Reference Manual". Online reference material. <http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html>, 2005.
9. Casella, George and Roger L. Berger. *Statistical Inference*. Brooks/Cole Publishing Company, Belmont, CA, 1990.
10. Cramér, Harald. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ, 1966.
11. Dempster, A.P., N.M. Laird, and D.B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society: Series B*, 39(1):1–21, November 1977.
12. Goldberger, Jacob, Shiri Gordon, and Hayit Greenspan. "An Efficient Image Similarity Measure based on Approximations of KL-Divergence Between Two Gaussian Mixtures". This unpublished paper was provided to the author by Jason Williams in Spring 2005.
13. Grewal, Mohinder S. and Angus P. Andrews. *Kalman Filtering: Theory and Practice*. John Wiley & Sons, New York, NY, 2001.
14. Herman, Shawn Michael. *A Particle Filtering Approach to Joint Passive Radar Tracking and Target Classification*. Ph.D. dissertation, Graduate College of the University of Illinois at Urbana-Champaign, Urbana, IL, April 2002.

15. Jebara, Tony and Risi Kondor. “Bhattacharyya and Expected Likelihood Kernels”. URL <http://www1.cs.columbia.edu/~jebara/papers/bhatta.pdf>. Columbia University, New York, NY, Unpublished Paper, 2003.
16. Julier, S.J. and J.K. Uhlmann. “A General Method for Approximating Nonlinear Transformations of Probability Distributions”. URL <http://citeseer.ist.psu.edu/julier96general.html>. Tech. Rep., RRG, Dept. of Engineering Science, University of Oxford, November 1996. <http://www.robots.ox.ac.uk/~siju/work/publications/letter size/Unscented.zip>.
17. Kullback, Solomon. *Information Theory and Statistics*. John Wiley & Sons, New York, NY, 1959.
18. Lainiotis, D.G. and S.K. Park. “On Joint Detection, Estimation and System Identification: Discrete Data Case”. *International Journal of Control*, 17(3):609–633, March 1973.
19. Lathi, B. P. *Modern Digital and Analog Communication Systems*. Oxford University Press, New York, NY, third edition, 1998.
20. Leon-Garcia, Alberto. *Probability and Random Processes for Electrical Engineering*. Addison-Wesley, Reading, MA, second edition, 1994.
21. Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, volume 1. Navtech, Arlington, VA, 1994.
22. Maybeck, Peter S. *Stochastic Models, Estimation, and Control*, volume 2. Navtech, Arlington, VA, 1994.
23. Maybeck, Peter S. “Multiple Model Adaptive Estimation”, 2005. Lecture Notes for Course EENG768. Air Force Institute of Technology, Wright-Patterson Air Force Base, OH.
24. Pollard, David. “Chapter 3: Distances and Affinities Between Measures”. URL <http://www.stat.yale.edu/~pollard/Asymptopia/>. Unpublished Chapter Excerpt from Pollard’s Draft Book, 17 October 2000.
25. Rao, C. Radhakrishna. *Linear Statistical Inference and Its Applications*. John Wiley & Sons, New York, NY, second edition, 1973.
26. Redner, Richard A. and Homer F. Walker. “Mixture Densities, Maximum Likelihood and the EM Algorithm”. *SIAM Review*, 26(2):195–239, April 1984.
27. Reid, Donald B. “An Algorithm for Tracking Multiple Targets”. *IEEE Transactions on Automatic Control*, AC-24(6):843–854, December 1979.
28. Ren, Cuirong, Dongchu Sun, and Dipak K. Dey. “Bayesian and Frequentist Estimation and Prediction for Exponential Distributions”. *Journal of Statistical Planning and Inference*, January 2005.
29. Ross, Sheldon M. *Introduction to Probability Models*. Academic Press, New York, NY, eighth edition, 2003.

30. Salmond, David J. "Mixture Reduction Algorithms for Target Tracking". *IEE Colloquium on State Estimation in Aerospace and Tracking Applications*, 7/1–7/4. IEE Publishing, London, UK, December 1989.
31. Salmond, David J. *Tracking in Uncertain Environments*. Technical Memorandum AW 121, Royal Aerospace Establishment, Farnborough, UK, September 1989. DTIC Number ADA215866. Taken from a D. Phil. thesis of the University of Sussex.
32. Singer, R.A., R.G. Sea, and K.B. Housewright. "Derivation and Evaluation of Improved Tracking Filters for use in Dense Multi-target Environments". *IEEE Transactions on Information Theory*, IT-20(4):423–832, July 1974.
33. Smith, Brian D. *Multiple Model Adaptive Estimator Target Tracker for Maneuvering Targets in Clutter*. Master's thesis, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, March 2005. AFIT/GE/ENG/05-18.
34. Strang, Gilbert. *Linear Algebra and its Applications*. Brooks/Cole Publishing Company, Belmont, CA, third edition, 1988.
35. Therrien, Charles W. *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall, Upper Saddle River, NJ, 1992.
36. Van Trees, Harry L. *Detection, Estimation, and Modulation Theory*, volume 1. John Wiley and Sons, Inc., New York, NY, 1968.
37. Weisstein, Eric W. "Hilbert Space". From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/HilbertSpace.html>.
38. Willams, Jason L. *Gaussian Mixture Reduction for Tracking Multiple Maneuvering Targets in Clutter*. Master's thesis, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, March 2003. AFIT/GE/ENG/03-19.
39. Williams, Jason L. SPIE Annual International Defense and Security Symposium, Orlando, Florida. Correspondence. April 2004.
40. Williams, Jason L. and Peter S. Maybeck. "Cost-Function-Based Gaussian Mixture Reduction for Target Tracking". *Proceedings of the Sixth International Conference of Information Fusion*, 1047–1054, Cairns, Australia, July 2003.
41. Williams, Jason L. and Peter S. Maybeck. "Cost-Function-Based Hypothesis Control Techniques for Multiple Hypothesis Tracking". *Proceedings of the SPIE Annual International Defense and Security Symposium*, 5428:167–179, Orlando, Florida, April 2004.
42. Zacks, Shelemyahu. *The Theory of Statistical Inference*. John Wiley & Sons, New York, NY, 1971.

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 22-12-2005			2. REPORT TYPE Master's Thesis		3. DATES COVERED (From — To) Sep 2004 — Nov 2005	
4. TITLE AND SUBTITLE Gaussian Mixture Reduction for Bayesian Target Tracking in Clutter			5a. CONTRACT NUMBER			
			5b. GRANT NUMBER			
			5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S) David J. Petrucci, Capt, USAF			5d. PROJECT NUMBER			
			5e. TASK NUMBER			
			5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management AFIT/EN 2950 Hobson Way WPAFB OH 45433-7765					8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/GE/ENG/06-01	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research Maj Todd E. Combs - (703) 696-9548 AFOSR/NM Suite 325, Room 3112 875 Randolph Street Arlington, Virginia 22203-1768					10. SPONSOR/MONITOR'S ACRONYM(S)	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approval for public release; distribution is unlimited.						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT The Bayesian solution for tracking a target in clutter results naturally in a target state Gaussian mixture probability density function (pdf) which is a sum of weighted Gaussian pdfs, or mixture components. As new tracking measurements are received, the number of mixture components increases without bound, and eventually a reduced-component approximation of the original Gaussian mixture pdf is necessary to evaluate the target state pdf efficiently while maintaining good tracking performance. Many approximation methods exist, but these methods are either <i>ad hoc</i> or use rather crude approximation techniques. Recent studies have shown that a measure-function-based mixture reduction algorithm (MRA) may be used to generate a high-quality reduced-component approximation to the original target state Gaussian mixture pdf. To date, the Integral Square Error (ISE) cost-function-based MRA has been shown to provide better tracking performance than any previously published Bayesian tracking in heavy clutter algorithm. Research conducted for this thesis has led to the development of a new measure function, the Correlation Measure (CM), which gauges the similarity between a full- and reduced-component Gaussian mixture pdf. This new measure function is implemented in an MRA and tested in a simulated scenario of a single target in heavy clutter. Results indicate that the CM MRA provides slightly better performance than the ISE cost-function-based MRA, but only by a small margin.						
15. SUBJECT TERMS Target Tracking, Multiple Hypothesis Tracking, Gaussian mixture reduction						
16. SECURITY CLASSIFICATION OF: a. REPORT U b. ABSTRACT U c. THIS PAGE U			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 189	19a. NAME OF RESPONSIBLE PERSON Dr. Peter S. Maybeck 19b. TELEPHONE NUMBER (include area code) (937) 255-3636, ext 4581	